

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 30, 2004, 14:30:23 ; Search time 10660.7 Seconds
(without alignments)
5378.881 Million cell updates/sec

Title: US-09-744-167-2

Perfect score: 6990

Sequence: 1 MWIDENAVEDQLIKRYSW.....QLSEGPVVMELIFYLENIV 1323

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Deiop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2 1/USPTO_epool_p/US09744167/runat_29042004_092919_23693/app_query.fasta_1.3214
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -PART=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09744167@cgn_1_16435@runat_29042004_092919_23693 -NCFU=6 -ICPU=3
-NO_MMAP -LARGESQRY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb_ba:*
2: gb_hug:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_hug_hum:*
31: em_hug_inv:*
32: em_hug_other:*
33: em_hug_mus:*
34: em_hug_pln:*
35: em_hug_rtd:*
36: em_hug_nam:*
37: em_hug_vrt:*
38: em_sv:*
39: em_higo_hum:*
40: em_higo_mus:*
41: em_higo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6937	99.2	4769	9	BC032680 Homo sapi
2	6937	99.2	4839	9	AF104304 Homo sapi
3	6551	93.7	4565	9	AF130419 Homo sapi
4	3978.5	56.9	4823	5	AF104305 Xenopus l
5	3497	50.0	2625	9	AF130420 Homo sapi
6	3397	48.6	1977	6	AX809313 Sequence
7	3333	47.7	133799	9	AC105754 Homo sapi
8	3333	47.7	150129	2	AC026909 Homo sapi
9	2623	37.5	89973	9	AL591398 Human DNA
10	2471.5	35.4	85802	10	AL671309 Mouse DNA
11	2368.5	33.9	258787	2	AC095988 Rattus no
12	2056	29.4	201330	2	AC118106 Rattus no
13	2047	29.3	5480	10	BC042669 Mus muscu
14	2045	29.3	5438	10	AX122241 Mus muscu
15	2042	29.2	6280	9	BSM805726 Homo sapi
16	2039	29.2	4620	9	AF434817 Homo sapi
17	2039	29.2	6632	9	AB002303 Human mRN
18	2020.5	28.9	5841	9	BSM804398 Homo sapi
19	1620.5	23.2	3793	9	AK127003 Homo sapi
20	1597.5	22.9	4598	3	AF239997 Drosophil
21	1597.5	22.9	4551	3	DME310804 Drosophil
22	1585.5	22.7	4660	3	AY051821 Drosophil
23	1528.5	21.9	888	6	AX677329 Sequence
24	1466.5	21.0	67833	3	AC004564 Drosophil
25	1466.5	21.0	179139	3	AC099307 Drosophil
26	1466.5	21.0	313634	3	AE003454 Drosophil
27	1462	20.9	54186	2	AC019884 Drosophil
28	1312	18.8	1237	10	BC046341 Mus muscu
29	1172	16.8	887	10	BC058416 Mus muscu
30	1122.5	16.1	4499	9	BC032227 Homo sapi
31	1046	15.0	1144	10	BC038255 Mus muscu
32	998	14.3	576	6	AX677675 Sequence
33	945	13.5	192533	9	AC055876 Homo sapi
34	941	13.5	151800	9	AC138649 Homo sapi
35	940	13.4	178257	2	AC024930 Homo sapi
36	940	13.4	180741	9	AC091565 Homo sapi
37	930.5	13.3	228470	2	AC136776 Homo sapi
38	816.5	11.6	5158	9	BSM805324 Homo sapi
39	812.5	11.6	1915	3	AK114777 Clona int
40	809	11.6	164706	9	AC068614 Homo sapi
41	569	8.1	4275	3	AF045926 Caenorhab
42	531	7.5	175593	2	BS111079 Danio rer
43	469.5	6.7	193289	2	AC127065 Rattus no
44	427.5	6.1	139101	2	BS111174 Danio rer
45	403.5	5.8	2902	9	BC030808 Homo sapi

ALIGNMENTS

RESULT 1

BC032680 4769 bp mRNA linear PRI 08-OCT-2003
LOCUS Homo sapiens MAD, mothers against decapentaplegic homolog
DEFINITION (Drosophila) interacting protein, receptor activation anchor, mRNA
(cDNA clone MGC:45259 IMAGE:5585817), complete cds.
ACCESSION BC032680.1 GI:21618607
VERSION MGC.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 4769)
Strausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, R.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schaeetz, R.E., Brownstein, M.J., Usdin, T.B., Tothiyuk, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shvchenko, Y., Bonfield, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzyzinski, M.I., Skalska, U., Smalios, D.E., Schnercher, A., Schein, J.E., Jones, S.J., Skalska, U., Marra, M.A., Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16906 (2002)
22388257
2 (bases 1 to 4769)
Strausberg, R.
Direct Submission
Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nigr.nih.gov
Akter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Grant, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, B., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tourgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
Series: IRAK Plate: 69 Row: p Column: 5
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4759059.
Location/Qualifiers
1..4769
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

gene
/clone="MGC:45259 IMAGE:5585817"
/tissue_type="Testis", embryonal carcinoma
/clone_lib="NIH_MGC_92"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..4769
/gene="MADHIP"
/note="synonyms: SARA, ZFYVE9"
/db_xref="LocusID:9372"
/db_xref="MIM:603755"
12..4289
/codon_start=1
/product="MADHIP protein"
/protein_id="AAH32680.1"
/db_xref="GI:21618608"
/db_xref="LocusID:9372"
/translation="MEN1;FOAENYLNKLVDEFQNEDETVSTILLDTKNKILDPDS
HRLSFTPIASVNESAVNSQPOLKVFSLAHSAPLTFEEDHCANODCNLNEIAT
XWIDENAEQLIKRYSWDDQCSAVEGKGCNGLACLPDEKNVLVAVMHNDXR
TLNDLQDNVNSQLMDAFSCSLDNENROTDOFSINESTKMMSEKQMDPLNR
PKTEGYSNHLCPSTSSDLSVCSQSLKDDGSGIRDFSMASITSLTVDSVSSQGT
GCPAYKAGNVIPEDLTGKISSPRTDLGSPNSFSHSEGLMKKEPAESTTESLR
SGLPULLKPMFNGSRNDCERCDCLFVNEVRADENEGHEBETLCTIEFLNMTX
FSESDMTNWLTKLNENNDQVNEKEFLQISPEDINGDSGCCVGLADAGLKH
CTCISECEDFTVIDTPAANY.SNGCDSYGMDPGVFPKTLPSKESVLSBEKEI
BESKECYNIQENAEATGSGILLNSTGLMKYLNHFCQVPSVLSQSSPKV
ASLPSISVFGGARPQPSNLKQIPKPLSHLQNDPANSNNNTKNDLIGAKLG
ENSTNVCSPLSGNISVDTNGHELESYEAEISTPCIALACGKGVFCASCSLACKLL
RKPFTILEPAPVWPDSQAQCMKCEAEFTTERHCRACGKGVFCASCSLACKLL
YNDREARVCVCHLVNMAQAENWMSASSQSPNPNPAEYCSITIPLOQAQASGL
SSPPTVMPVGVKHPGAENQREQRVWFADCLPNEGVAADAKLMTGTSAGT
LAVSDPVKPTISPLPAETDCLFSGISTQVSPGSANLIPEDGLPILISTGVK
GDYAVEKPSQISVMQLEDGDPDLFVFLNANLLSMVKIYVYVNRKWCFTTGMEA
VQSPBIVTLOCLPDKCLPKDIFNHVQLYQVLDALAGNVSNLGHLSFSGFLSGKH
GGFLVSTYQSLQDLVLPDLYLFGILIQWETPAKVPFIRLMRLGASRYLPQP
LFSVFRPLRFGEGHTIMLLADPRVQVTLPUVQGLVDMVEYKTSIKTPSNRYE
MKKNNKNEHVLAGGACFNEKADSHLVQVNDGNYQVQTAISIHNPRTKVTGASFTV
FSGALKSSGYLAKSIVEDGMVQITAEENDSLRQALREMKDFTTCGKADAEPOE
HIHIQWDDDNKSVGVSPIDGKSMETITNVI.FHGEYKANGKVRITWTEVFLEND
DOHNCPLPADHSRLTEHVAKFCLALCPHLKLEDGMLKGLRVLDSQVGYQAG
SNGQLPSPQYMNLDLSALVPVTHGACQLSGPVMVMEILFYLEINIV"
2091..2282
/note="FYVE, Region: FYVE zinc finger. The FYVE zinc
finger is named after four proteins that it has been found
in: FABI, YOTB/ZK632.12, Vaci1, and EEA1. The FYVE finger
has been shown to bind two Zn++ ions. The FYVE finger has
eight potential zinc coordinating cysteine positions. Many
members of this family also include two histidines in a
motif R-HHC-XCG, where + represents a charged residue and
X any residue. We have included members which do not
conserve these histidine residues but are clearly related"
/db_xref="CDD:pfam01363"

misc_feature
US-09-744-167-2 (1-1323) x BC032680 (1-4769)
QY 1 MetTrpIleAspGluAsnAlaValIaGluAspGlnLeuIleLysArgAsnTySerTrp 20
Db 318 ATGTGGATTGATGAAATGCTGTTGAGAGACCACTTAATTAAGAGAACTATAGTTGG 377
QY 21 AspAspGlnCysSerAlaValGluValGlyGluLysCysGlyAsnLeuAlaCysLeu 40
Db 378 GATGATCAATGACGTGCTGTTGAGTGGAGAGAAATGTGGAAACCTGGCTGTCTG 437
QY 41 ProAspGluLysAsnValLeuValValAlaValMetHisAsnCysAspLysArgThrLeu 60

ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 4769
Score: 6337.00 Matches: 1315
Percent Similarity: 99.47% Conservative: 1
Best Local Similarity: 99.40% Mismatches: 7
Query Match: 99.24% Indels: 0
DB: 9 Gaps: 0

Db 438 CCAGATGAGAGAAATGTTCTTTGTTGTTAGCCCTCATGCATAAATGTTGATAAAAGCAATTA 497
Qy 61 GlnAsnAspLeuGlnAspCysAsnAsnTyrAsnSerGlnSerLeuMetAspAlaPheSer 80
Db 498 CAAACCAATTTACAGGATTGTAATTAATTAATAGTCAATCCCTTTATGGATGCTTTTACG 557
Qy 81 CysSerLeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGluSer 100
Db 558 TGTTCACTGGATATAAGAAACAGACAAATGATCAATTTAGTTTATAGTATAAATGAGTCC 617
Qy 101 ThrGluLysAspMetAsnSerGluLysGlnMetAspProLeuAsnArgProLysThrGlu 120
Db 618 ACTGAAAAAGATATGAATTCAGAGAAACAAATGGATCCATTTGAATGACCGAAACAGAG 677
Qy 121 GlyArgSerValAsnHisLeuCysProThrSerSerAspSerLeuAlaSerValCysSer 140
Db 678 GGGAGATCTGTTAACCATCTGTGCTTACCTTCATCTCATAGTCTAGCCAGTGTCTGTTC 737
Qy 141 ProSerGlnLeuLysAspAspGlySerIleGlyArgAspProSerMetSerAlaIleThr 160
Db 738 CCTTCACAATTAAGAGATGAGGAGATATAGGTAGAGACCCCTCCATGTCTGGCATTA 797
Qy 161 SerLeuThrValAspSerValIleSerSerGlnGlyThrAspGlyCysProAlaValLys 180
Db 798 AGTTTACGGTTGATTCAGTAATCTCAATCCAGGAAACAGATGATGCTCTGTGTTAA 857
Qy 181 LysGlnGluAsnTyrIleProAspGluAspLeuThrGlyLysIleSerSerProArgThr 200
Db 858 AAGCAAGAGAACTATATACAGATGAGGACCTCACTGGCAAAATCAGCTCTCTAGGACA 917
Qy 201 AspLeuGlySerProAsnSerPheSerHisMetSerGluGlyIleLeuMetLysLysGlu 220
Db 918 GATCTAGGAGTCCAAATTCCTTTCCCATGATGAGGAGATTTTGATGAAAAAGAG 977
Qy 221 ProAlaGluGluSerThrGluGluSerLeuArgSerGlyLeuProLeuLeuLeuLys 240
Db 978 CCAGCAGAGAGACCACTGAGATCCCTCCGGTCTGGTTTACCTTTGCTTCTCATA 1037
Qy 241 ProAspMetProAsnGlySerGlyArgAsnAsnAspCysGluArgCysSerAspCysLeu 260
Db 1038 CCAGACATGCTATATGGTCTGGAAGGAATTAATGACTGTGAACGGTGTTCAGATTGCC 1097
Qy 261 ValProAsnGluValArgAlaAspGluAsnGluGlyTyrGluHisGluGluThrLeuGly 280
Db 1098 GTGCCTTAATGAGTTAGGCTGATGAAATGAAGTTATGAATGAAGAAATCTTTGGC 1157
Qy 281 ThrThrGluPheLeuAsnMetThrGluHisPheSerGluSerGlnAspMetThrAsnTyr 300
Db 1158 ACTACAGAAATCCTTAATATGACAGACATTTCTCTGAATCTCAGGACATGACTAATGG 1217
Qy 301 LysLeuThrLysLeuAsnGluMetAsnAspSerGlnValAsnGluGluLysPhe 320
Db 1218 AAGTTGACTAACTAAATGAGATGATGATAGCCAAATGAACGAAAGAAAGAAAGTTT 1277
Qy 321 LeuGlnIleSerGlnProGluAspThrAsnGlyAspSerGlyGlyLncysValGlyLeu 340
Db 1278 CTACAGATTAGTCAGCTCAGGACACTAATGGTGTATGATGAGGACAGTGTGTGATTG 1337
Qy 341 AlaAspAlaGlyLeuAspLeuLysGlyThrCysIleSerGluSerGluGluCysAspPhe 360
Db 1338 GCAGATCAGGCTTAGATTTTAAAGGAACCTTGATTTAGTGAAGTGAAGAAATGTGATTTC 1397
Qy 361 SerThrValIleAspThrProAlaAlaAsnTyrLeuSerAsnGlyCysAspSerTyrGly 380
Db 1398 TCCACTGTTATGACACACAGGAGCANNATATCTATCTAATGTTGTGATTCCTATGGA 1457
Qy 381 MetGlnAspProGlyValSerPheValProLysThrLeuProSerLysGluAspSerVal 400
Db 1458 ATGCAAGACCCAGGTGTTCTTTTGTTCCTCAAGACCTTTACCTCCAAAGAGATTTCAGTA 1517
Qy 401 ThrGluLysGluIleGluGluSerLysSerGluCysTyrSerAsnIleTyrGluGln 420

Db 1518 ACAGAGAAAAAGAAATAGAGAAAGCAAGTCAAGATGCTCAAAATATTTATGAACAG 1577
Qy 421 ArgGlyAsnGluAlaThrGluGlySerGlyLeuLeuLeuAsnSerThrGlyAspLeuMet 440
Db 1578 AGAGGAATAGAGCCACAGAGAGGAGTGACACTACTTTTAAACAGCACTGGTCACTTAATG 1637
Qy 441 LysLysAsnTyrLeuHisAsnPheCysSerGlnValProSerValLeuGlyGlnSerSer 460
Db 1638 AAGAAAAATTTATCATAAATTTCTGTAGTCAAGTTCATCAGTCTCTGGGCAATCTTCC 1697
Qy 461 ProLysValValAlaSerLeuProSerIleSerValProPheGlyGlyAlaArgProLys 480
Db 1698 CCCAAGGTAGTAGCAAGCTGCCATCTATCAGTGTCTCTTTTGGTGGTGCAAGCCCAAG 1757
Qy 481 GlnProSerAsnLeuLysLeuGlnIleProLysProLeuSerAspHisLeuGlnAsnAsp 500
Db 1758 CAACCTTCTAATCTTAACTTCAAATCCAAAGCCATTATCAGACCAATTTACAAATGAC 1817
Qy 501 PheProAlaAsnSerGlyAsnAsnThrLysAsnLysAsnAspIleLeuGlyLysAlaLys 520
Db 1818 TTTCTCTGCAACAGTGGAAATAATACTAAAAATAAAAAATGATATTCTTTGGGAAACAAA 1877
Qy 521 LeuGlyGluAsnSerAlaThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal 540
Db 1878 TTAGGGGAAACTCAGCAACCAATGTATGCAGTCCATCTTTGGGAAACATCTCTAATGTC 1937
Qy 541 AspThrAsnGlyGluHisLeuGluSerTyrGluAlaGluIleSerThrArgProCysLeu 560
Db 1938 GATACAAATGGGAAACATTTAGAAAGTTATGAGGCTGAGATCTCCACTAGACCATGCTT 1997
Qy 561 AlaLeuAlaProAspSerProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAla 580
Db 1998 GATTAGCTCCAGATAGCCAGCAATATGATCTCAGAGCTGGTCAGTTGGAAATTTCTGCC 2057
Qy 581 ArgLysProPheThrThrLeuGlyGluValAlaProValTyrValProAspSerGlnAla 600
Db 2058 AGAAAGCCATTCAACACTCTGGTCTGAGTGGTCCAGTATGGTACCGGATTTCTCAGCT 2117
Qy 601 ProAsnCysMetLysCysGluAlaArgPheThrPheThrLysArgArgHisCysArg 620
Db 2118 CCAAAATTCATGAAATGTGAAGCCAGGTTTACATTCACCAAAAGAGGACATCAGTCAGA 2177
Qy 621 AlaCysGlyLysValPheCysAlaSerCysCysSerLeuLysCysLysLeuLeuTyrMet 640
Db 2178 GCATGTGGAGGTTTCTGTGCTTCTCTGTAGCTGAAATGTAAATGTTTATACATG 2237
Qy 641 AspArgLysGluAlaArgValCysValIleCysHisSerValLeuMetAsnAlaGlnAla 660
Db 2238 GACAGAAAGGAAGCTAGAGTGTGTAAATCTGCCATTCACTGCTCAATGAATGCTCAAGCC 2297
Qy 661 TyrGluAsnMetMetSerAlaSerSerGlnSerProAsnProAsnAsnProAlaGluTyr 680
Db 2298 TGGGAGACATGATGAGTGCCTCAAGCCAGAGCCCTAACCCCTAACCAATCTCTGCTGAATAC 2357
Qy 681 CysSerThrIleProProLeuGlnGlnAlaGlnAlaSerGlyAlaLeuSerSerProPro 700
Db 2358 TGTTCTACTATCCCTCCCTTGAGCAAGCTCAGGCCCTCAGGAGCTCTCAGCTCTCCACT 2417
Qy 701 ProThrValMetValProValGlyValLeuLysHisProGlyAlaGluValAlaGlnPro 720
Db 2418 CCCACTGTGATGGTACCTGTGGAGTTTAAAGCACCCCTGGAGCAGAAAGTGGCTCAGGCC 2477
Qy 721 ArgGluGlnArgArgValTyrPheAlaAspGlyIleLeuProAsnGlyGluValAlaAsp 740
Db 2478 AGAGAGCAGAGCGGAGTTTGGTTTGTGATGGGATCTTGCCCAATGGAGAGTTGCTGAT 2537
Qy 741 AlaAlaLysLeuThrMetAsnGlyThrSerSerAlaGlyThrLeuAlaValSerHisAsp 760
Db 2538 GCAGCCAAATTAACAATGAATGGAATCTCTCTGCAGGAACCCCTGGCTGTGTACACAGAC 2597
Qy 761 ProValLysProValThrThrSerProLeuProAlaGluThrAspIleCysLeuPheSer 780
Db 2598 CCAGTCAAGCCAGTAACTACACAGTCTCTACAGCAGACAGAGATATTGTTCTATTCTCT 2657

781 GlySerIleThrGlnValGlySerProValGlySerAlaMetAsnLeuIleProGluAsp 800
 Db GGGAGTATAACTCAGGTTGGAAGTCTGTTGGAAGTGCATGAATCTTATTCCTGAAGAT 2717
 801 GlyLeuProProIleLeuIleSerThrGlyValGlyGlyAspTyrAlaValGluGluLys 820
 Db GGCCTTCCTCCCATCTCACTCCACTGGGTGTAAGAGACATGCTGTGGGAAGAA 2777
 821 ProSerGlnIleSerValMetGlnGlnLeuGluAspGlyGlyProAspProLeuValPhe 840
 Db CCATCAGAGATTCAGTAATGCACAGCTTGAGGATGGTGGCCCTGACCCACCTGTATTT 2837
 841 ValLeuAsnAlaAsnLeuLeuSerMetValLysIleValAsnTyrValAsnArgLysCys 860
 Db GTTTTAAATGCAAAATTTGTTGTCATGTTGTTAAATTTGTAATTTATGTGAACAGGAAGTGC 2897
 861 TrpCysPheThrThrLysGlyMetHisAlaValGlyGlnSerGluIleValIleLeuLeu 880
 Db TGGTGTTCACAAACCAAGGGAATGCATGCGAGTGGTCACTCGAGATAGTCACTTCCTCA 2957
 881 GlnCysLeuProAspGluLysCysLeuProLysAspIlePheAsnHisPheValGlnLeu 900
 Db CAGTGTTCACCGGATGAAAGTGTTCGCAAGGATATCTTTAATCACTTTGTGCAGCTT 3017
 901 TyrArgAspAlaLeuAlaGlyAsnValValSerAsnLeuGlyHisSerPhePheSerGln 920
 Db TATCGGATGCTCTGGCAGGGAATGTGTGAGCACTTGGGACATCTCTTCCTTCAGTCAA 3077
 921 SerPheLeuGlySerLysGluHisGlyGlyPheLeuTyrValThrSerThrTyrGlnSer 940
 Db AGTTTCCTTGGCAGTAAAGAACATGGTGGATCTTATATGTGACATCTACCTACCACTCA 3137
 941 LeuGlnAspLeuValLeuProThrProProTyrLeuPheGlyIleLeuIleGlnLysTrp 960
 Db CTGCAAGACCTTAGTACCTCCCAACCCACCTTACTTGTGTGGGATCTTATCCGAAGTGG 3197
 961 GluThrProTyrAlaLysValPheProIleArgLeuMetLeuArgLeuGlyValaGluTyr 980
 Db GAACCTCTTGGGTAAAGTATTTCTTATCCGTCTGATGTGAGACTTGGAGCTGAATAT 3257
 981 ArgLeuTyrProCysProLeuPheSerValArgPheArgLysProLeuPheGlyGluThr 1000
 Db CGACTTTATCATGCCCATCTTATTCAGTCTCAGATTTCCGAAGCCATGTTTGGAGAGAC 3317
 1001 GlyHisThrIleMetAsnLeuAlaAspPheArgAsnTyrGlnTyrThrLeuProVal 1020
 Db GGGCATACCATCATGAATCTTCTTGAGACTTCAGAAATACCAAGTATACCTTGCCAGTA 3377
 1021 ValGlnGlyLeuValValAspMetGluValArgLysThrSerLeuLysIleProSerAsn 1040
 Db GTTCAAGGTTTGGTGGTGTATGATGAAAGTTCGGAAGTCTAGCATCAAAATTTCCAGCAAC 3437
 1041 ArgTyrAsnGluMetMetLysAlaMetAsnLysSerAsnGluHisValLeuAlaGlyGly 1060
 Db AGATACATGATGATGATGAAGCCATGAACAAAGTCCATGAGCATGTCCTGGCAGGAGGT 3497
 1061 AlaCysPheAsnGluLysAlaAspSerHisLeuValCysValGlnAsnAspGlyAsn 1080
 Db GCCTGCTTCAATGAAAGGCGACACTCTCATCTTGTGTGTACAGAAATGATGATGAAAC 3557
 1081 TyrGlnThrGlnAlaIleSerIleHisAsnGlnProArgLysValThrGlyValAspPhe 1100
 Db TATCAGACCCAGGATCATCAGTATTCAATCAGCCAGAAAGATGATGTCCTGTCAGTTTC 3617
 1101 PheValPheSerGlyValaLeuLysSerSerGlyTyrLeuAlaLysSerSerIleVal 1120
 Db TTTGTGTTCACTGGCGCTCTCAATCTCTCTGATACCTTGCCTCAAGTCCAGTATTGTG 3677
 1121 GluAspGlyValMetValGlnIleThrAlaGluAsnMetAspSerLeuArgGlnAlaLeu 1140
 Db GAAGATGGTGTATGGTCCAGATTACTGCAGAAACATGGATTCTCTTGGAGCGAGGCACTG 3737

1141 ArgGluMetLysAspPheThrIleThrCysGlyLysAlaAspAlaGluGluProGlnGlu 1160
 Db CGAGAGATGAAGGACTTCACCATCACCTGTGGAGGCGGACGGAGGAGAACCCAGAG 3797
 1161 HisIleHisIleGlnTyrValAspAspLysAsnValSerLysGlyValValSerPro 1180
 Db CACATCCACATCCAGTGGTGGATGATGACAAAGACGTTAGCAAGGCTGCTAAGTCT 3857
 1181 IleAspGlyLysSerMetGluThrIleThrAsnValLysIlePheHisGlySerGluTyr 1200
 Db ATAGATGGGAAGTCCATGGAGACTATACAAATGTGAAGATATTCCATGGATCAGAAAT 3917
 1201 LysAlaAsnGlyLysValIleArgTyrThrGluValPhePheLeuGluAsnAspGln 1220
 Db AAGCAATGGAAGATTAATCAGATGGACAGAGTGTCTTTCTAGAAAACGATGACCA 3977
 1221 HisAsnCysLeuSerAspProAlaAspHisSerArgLeuThrGluHisValaLysAla 1240
 Db CACAATTCCTCAGTATCTCTGCAGATCACAGTATGATGAGCATGTTCGCAAGCT 4037
 1241 PheCysLeuAlaLeuCysThrGlnLeuLysLeuLysGlyAspGlyMetThrLysLeu 1260
 Db TTTTCCCTTGTCTCTCTCTCTCCTCCTGAACTTCTGAGGAGATGGATGACCAACTG 4097
 1261 GlyLeuArgValThrLeuAspSerAspGlnValGlyTyrGlnAlaGlySerAsnGlyGln 1280
 Db GGACTACCTGTGACACTTGTGACTCAGATCAGTGGTGTGATCAAGCAGGAGCAATGGCC 4157
 1281 HisLeuProSerGlnTyrMetAsnAspPheAspSerAspLeuValLysMetIleHisGly 1300
 Db CCCCTTCCCTCGCAGTATCAATGATCTGACAGCGCTTGGTCCGGTATCATGGA 4217
 1301 GlyAlaCysGlnLeuSerGluGlyProValValMetGluLeuIlePheTyrIleLeuGlu 1320
 Db GGGGCTCCAGCTTAGTAGGCGCCCGTGTCTGATGAACTCATCTTTATATTCTGAA 4277
 1321 AsnIleVal 1323
 Db AACATCGTA 4286

RESULT 2
 AF104304 4839 bp mRNA linear PRI 03-JAN-1999
 LOCUS Homo sapiens Smad anchor for receptor activation (SARA) mRNA,
 complete cds.
 ACCESSION AF104304
 VERSION AF104304.1 GI:4092766
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 4839)
 AUTHORS Tsukazaki,T., Chiang,T.A., Davison,A.F., Attisano,L. and Wrana,J.L.
 TITLE SARA, a FIVE domain protein that recruits Smad2 to the TGFbeta
 RECEPTOR
 JOURNAL Cell 95 (6), 779-791 (1998)
 MEDLINE 99081294
 PUBMED 9865696
 REFERENCE 2 (bases 1 to 4839)
 AUTHORS Wrana,J.L.
 TITLE Direct Submission
 JOURNAL Submitted (04-NOV-1998) Program in Developmental Biology, Hospital
 for Sick Children, 555 University Avenue, Toronto, Ontario M6G 1X8,
 Canada

FEATURES
 source
 1..4839
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /cell_type="HepG2"
 1..4839
 /gene="SARA"

gene

QY 521 LeuGlyGluAsnSerAlaThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal 540
Db 1999 ITAGGGGAAACTCAGCAACCAATGATGAGTCCATCTTTGGGAAACATCTCTAATGTC 2058
QY 541 AspThrAsnGlyGluHisLeuGluSerTyGluAlaGluIleSerThrArgProCysLeu 560
Db 2059 GATCAAAATGGGAACATTTAGAAGATTATGAGGCTGAGATCTCCACTAGACCATGCCCT 2118
QY 561 AlaLeuAlaProAspSerProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAla 580
Db 2119 GCATTAGCTCAGATAGCCAGATAATGATCTCAGAGCTGGTCAGTTTGGAAATTTCTGCC 2178
QY 581 ArgIysProPheThrLeuGlyGluValAlaProValTrpValProAspSerGlnAla 600
Db 2179 AGAAAGCCATTTCACACGCTGGGTGAGTGCTCCAGTATGGGTACCGGATCTCTCAGGCT 2238
QY 601 ProAsnCysMetLysCysGluAlaArgPheThrPheThrLysArgArgHisCysArg 620
Db 2239 CCAAAATTCATGAATGTGAAGCCAGGTTTACATTCACCAAAAGGAGGCACTCACTGCAGA 2298
QY 621 AlaCysGlyLysValPheCysAlaSerCysCysSerLeuLysCysLysLeuLeuTyrMet 640
Db 2299 GCATGTGGGAAGTTTCTGTGCTTCTGCTGAGCTGAAATGTAAACTGTTATACATG 2358
QY 641 AspArgIysGluAlaArgValCysValIleCysHisSerValLeuMetAsnAlaGlnAla 660
Db 2359 GACAAAGGAAGCTAGAGTGTGTAACTTGCCATCTCAGTCTGATGAATGCTCAAGCC 2418
QY 661 TrpGluAsnMetMetSerAlaSerGlnSerProAsnProAsnProAlaGluTyr 680
Db 2419 TGGGAGAACATGATGAGTGCCTCAAGCCAGAGCCCTTAACCTTAACATCTGCTGAATAC 2478
QY 681 CysSerThrIleProProLeuGlnGlnAlaGlnIleAsnSerGlyAlaLeuSerSerProPro 700
Db 2479 TGTTCTACTATCTCCCTCCCTGACAGCAAGCTCAGCCCTCAGGAGCTCTGAGCTCTCCACCT 2538
QY 701 ProThrValMetValProValGlyValLeuLysHisProGlyAlaGluValAlaGlnPro 720
Db 2539 CCCACTGTGATGGTACCTGTGGAGTTTAAAGCACCTCGAGCAGAAAGTGGCTCAGCCC 2598
QY 721 ArgGluGlnArgArgValTrpPheAlaAspGlyIleLeuProAsnGlyGluValAlaAsp 740
Db 2599 AGAGAGCAGAGCGGAGTTTGGTTTGTGATGGGATCTTGCCCAATGGAGAAAGTGTCTGAT 2658
QY 741 AlaAlaLysLeuThrMetAsnGlyThrSerSerAlaGlyThrLeuAlaValSerHisAsp 760
Db 2659 GCAGCCAAATTAACAATGAATGGAACTTCTCTCGAGAACCCCTGGCTGTGTACACGAC 2718
QY 761 ProValLysProValThrThrSerProLeuProAlaGluThrAspIleCysLeuPheSer 780
Db 2719 CCAGTCAAGCCAGTAACCTACCAGTCTCTACAGCAGAGCGGATATTTGCTATTCTCT 2778
QY 781 GlySerIleThrGlnValGlySerProValGlySerAlaMetAsnLeuIleProGluAsp 800
Db 2779 GGGAGTAACTCAGGTTGGAAGTCTGTTGGAAGTGAATGAATCTATTCTCTGAAGAT 2838
QY 801 GlyLeuProPheIleLeuIleSerThrGlyValLysGlyAspTyAlaValGluGluLys 820
Db 2839 GGCCTTCTCCATCTCATCTCCACTGGTGTAAAGAGGAGCTATGCTGTGGAAGAGAA 2898
QY 821 ProSerGlnIleSerValMetGlnGlnLeuGluAspGlyGlyProAspProLeuValPhe 840
Db 2899 CCATCAGAGATTTTCAGTAATGACAGCTGGAGGATGGTGGCCCTGACCCACTGTATTT 2958
QY 841 ValLeuAsnAlaAsnLeuSerMetValLysIleValAsnTyValAsnArgLysCys 860
Db 2959 GTTTTAATGCAATTTGTTGTAATGGTAAATTTGTAATTTATGTGAACAGAGAGTGC 3018
QY 861 TrpCysPheThrThrLysGlyMetHisAlaValGlyGlnSerGluIleValIleLeuLeu 880
Db 3019 TGGTGTCTTCAACAAAGGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3078
QY 881 GlnCysLeuProAspGluLysCysLeuProLysAspIlePheAsnHisPheValGlnLeu 900

Db 3079 CAGTGTCTTACCGGATGAAAAGTGTTCCAAAGGATATCTTTAATCACTTTGTGACGCTT 3138
QY 901 TyrArgAspAlaLeuAlaGlyAsnValValSerAsnLeuGlyHisSerPhePheSerGln 920
Db 3139 TATCGGGATGCTTGGCAGGGAATGTGTGAGCAACTTGGGACATTCCTTCTTCAGTCAA 3198
QY 921 SerPheLeuGlySerLysGluHisGlyGlyPheLeuTyrValThrSerThrTyGlnSer 940
Db 3199 AGTTTCTTGGCAGTAAAGAAACATGGTGGATCTTATATGTGACATCTACCTACAGTCA 3258
QY 941 LeuGlnAspLeuValLeuProThrProTyrLeuPheGlyIleLeuIleGlnLysTrp 960
Db 3259 CTGCAAGACCTAGTACTCCCAACCCCACTTACTTGTGGGATCTTATCCAGAAATGG 3318
QY 961 GluThrProTrpAlaLysValPheProIleArgIleuMetLeuArgLeuGlyAlaGluTyr 980
Db 3319 GAAACTCTCTGGGCTAAAGTATTTCTTATCGCTCTGATGTTGAGACTTGGAGCTGAATAT 3378
QY 981 ArgLeuTyrProCysProLeuPheSerValArgPheArgLysProLeuPheGlyGluThr 1000
Db 3379 CGACTTTATCCATGCCCACTATTCACTGTGAGTTTCGAAAGCCATTTCTTTGGAGAGCG 3438
QY 1001 GlyHisThrIleMetAsnLeuAlaAspPheArgAsnTyrGlnTyrThrLeuProVal 1020
Db 3439 GGGCATACCATCATGAATCTTCTTCAGACTTCAGAAATTTACCAATATACCATGATATACCTGCCAGTA 3498
QY 1021 ValGlnGlyLeuValValAspMetGluValArgLysThrSerIleLysIleProSerAsn 1040
Db 3499 GTTCAAGTGTGGTGGTGTGATGGAAGTTCGGAATCTAGCAAAATTTCCAGCAAC 3558
QY 1041 ArgTyrAsnGluMetMetLysAlaMetAsnLysSerAsnGluHisValLeuAlaGly 1060
Db 3559 AGATCAATGAGATGATGAAAGCCATGAACAGTCCAATGAGCATGCTCTGCAGGAGGT 3618
QY 1061 AlaCysPheAsnGluLysAlaAspSerHisLeuValCysValGlnAsnAspAspGlyAsn 1080
Db 3619 GCTCTCTTCAATGAAAGGCGAGACTCTCATCTGTGTGTGTACAGAATGATGATGGAAC 3678
QY 1081 TyrGlnThrGlnAlaIleSerIleHisAsnGlnProArgLysValThrGlyAlaSerPhe 1100
Db 3679 TATCAGACCCAGCTTATCAGTATTTCACATCAGCCAGAAAGTACTGCTGCTCCAGTTTC 3738
QY 1101 PheValPheSerGlyAlaLeuLysSerSerGlyTyrLeuAlaLysSerSerIleVal 1120
Db 3739 TTTGTGTTCAGTGGGCTCTGAAATCTCTTCTGATACCTTGCACAGTCCAGTATTGTG 3798
QY 1121 GluAspGlyValMetValGlnIleThrAlaGluAsnMetAspSerLeuArgGlnAlaLeu 1140
Db 3799 GAAGATGTGTGTATGTTCCAGATTAATCAGAGCAACATGGATTCCTTGGCAGGACTG 3858
QY 1141 ArgGluMetLysAspPheThrIleThrCysGlyLysAlaAspAlaGluGluProGlnGlu 1160
Db 3859 CGAGAGATGAAGACTTCACCATCACCTGTGGGAAGCGGACGCGAGAAACCCAGAG 3918
QY 1161 HisIleHisIleGluTrpValAspAspLysAsnValSerLysGlyValValSerPro 1180
Db 3919 CACATCCCATCCAGTGGGTGATGATCAAGAACCTTAGCAAGGGTGTCTGTAAGTCTCT 3978
QY 1181 IleAspGlyLysSerMetGluThrIleThrAsnValLysIlePheHisGlySerGluTyr 1200
Db 3979 ATAGATGGGAAGTCCATCGAGACTATAACAATGTGAAGATATTCATGGATCAGAAATAT 4038
QY 1201 LysAlaAsnGlyLysValIleArgTrpThrGluValPhePheLeuGluAsnAspGln 1220
Db 4039 AAAGCAAAATGCAAAAGTAAATCAGATGGACAGAGGTGTTTTTCTTAGAAACCATGACAG 4098
QY 1221 HisAsnCysLeuSerAspProAlaAspHisSerArgLeuThrGluHisValAlaLysAla 1240
Db 4099 CCAATTTGCTCAGTGTATCTTCAGATCAGATGATGATGATGATGATGATGATGATGAT 4158
QY 1241 PheCysLeuAlaLeuCysThrGlnLeuLysLeuLysGlyAspGlyMetThrLysLeu 1260

Db 1139 CCAGCAGAGGAGCACCACCTGAAAGAAATCCCTCGGTCTGGTTTACCTTTGCTTCTCAA 1198
Qy 241 ProAspMetProAsnGlySerGlyArgAsnAsnAspCysGluArgCysSerAspCysLeu 260
Db 1199 CCAGACATGCTAATGGTCTGGAGGAATTAAGTCTGTGAACGGTGTTCAGATTGCCIT 1258
Qy 261 ValProAsnGluValArgAlaAspGluAsnGluGlyTyrGluHisGluGluThrLeuGly 280
Db 1259 GTGCCTAATGAAGTTAGGGCTGATGAAATCAAGGTTATGAACATGAAGAACTCTTTGC 1318
Qy 281 ThrThrGluPheLeuAsnMetThrGluHisPheSerGluSerGlnAspMetThrAsnTrp 300
Db 1319 ACTACAGAAATTCCTTAATATGACAGACATTTCTGAAATCTCAGACATGACTAATGG 1378
Qy 301 LysLeuThrLysLeuAsnGluMetAsnAspSerGlnValAsnGluGluLysGluLysPhe 320
Db 1379 AAGTTGACTAACTAATGAGATGAATGATAGCAAGTAAACGAAGAAAGAAAGTAT 1438
Qy 321 LeuGlnIleSerGlnProGluAspThrAsnGlyAspSerGlyGlnCysValGlyLeu 340
Db 1439 CTACAGATTAGTCAGCCTGAGGACACTAATGGTATAGTGGAGACAGTGTGTGGATTG 1498
Qy 341 AlaAspAlaGlyLeuAspLeuLysGlyThrCysIleSerGluSerGluGluCysAspPhe 360
Db 1499 GCAGATCGAGTCTAGATTTAAAGGAACITTCATTAGTAGTGAAGTGAAGATGTGATTC 1558
Qy 361 SerThrValIleAspThrProAlaAlaAsnTyrLeuSerAsnGlyCysAspSerTyrGly 380
Db 1559 TCCACTGTTATAGACACACACAGCAGCAATATCTATCTAATGGTGTGTGATTCCTATGGA 1618
Qy 381 MetGlnAspProGlyValSerPheValProLysThrLeuProSerLysGluAspSerVal 400
Db 1619 ATGCAGACCCAGGTGTTCTTTGTTCCAAAGACTTACCTCCAAAGAGATTCACTA 1678
Qy 401 ThrGluGluLysGluIleGluGluSerLysSerGluCysTyrSerAsnIleTyrGluGln 420
Db 1679 ACAGAGAAAGAAATAGAGAAAGCAAGTCAGATGCTACTCAAAATATTTATGAACAG 1738
Qy 421 ArgGlyAsnGluAlaThrGluGlySerGlyLeuLeuAsnSerThrGlyAspLeuMet 440
Db 1739 AGAGGAATAGGCCACAGAGGAGTGGACTCTTTTAAACAGCAGCTGTGTGACCTAATG 1798
Qy 441 LysLysAsnTyrLeuHisAsnPheCysSerGlnValProSerValLeuGlyGlnSerSer 460
Db 1799 AAGAAATATTTATACATAATTTCTGTAGTCAAGTTCATCAGTGTCTGGCAATCTTCC 1858
Qy 461 ProLysValValAlaSerLeuProSerIleSerValProPheGlyAlaArgProLys 480
Db 1859 CCCAAGGTAGTAGCAAGCCTGCCATCTATCAGTGTTCCTTTGGTGTGCAAGCCCAAG 1918
Qy 481 GlnProSerAsnLeuLysLeuGlnIleProLysProLeuSerAspHisLeuGlnAsnAsp 500
Db 1919 CAACCTCTAATCTTAACTTCAATTCGAAGCCATTATCAGACCATTTACAAATGAC 1978
Qy 501 PheProAlaAsnSerGlyAsnAsnThrLysAsnLysAsnAspIleLeuGlyLysAlaLys 520
Db 1979 TTTCTCTGAAACAGTGGAAATAATACTAAAAATAAAATGATATCTTTGGGAAAGCAAAA 2038
Qy 521 LeuGlyGluAsnSerAlaThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal 540
Db 2039 TTAGGGMAAACTCAGCAACCAATGTATGAGTCCATCTTTGGGAAACATCTCTAATGTC 2098
Qy 541 AspThrAsnGlyGluHisLeuGluSerTyrGluAlaGluIleSerThrArgProCysLeu 560
Db 2099 GATCAAAATGGGAACATTTAGAAAGTTATGAGGCTGAGATCTCCACTAGACCATGCTT 2158
Qy 561 AlaLeuAlaProAsnSerProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAla 580
Db 2159 GCATTAGCTCAGATAGCCCAAGTAATGATCTCAGACCTGTGATTTGGAATTTCTGCC 2218
Qy 581 ArgLysProPheThrThrLeuGlyGluValAlaProValTrpValProAspSerGlnAla 600
Db 2219 AGAAAGCAATTCACCACTCTGGGTGAGGTGGCTCCAGTATGGGTACCGGATTTCTCAGGCT 2278

Qy 601 ProAsnCysMetLysCysGluAlaArgPheThrPheThrLysArgHisHisCysArg 620
Db 2279 CCAAATTCATGAATGTGAAGCCAGGTTTACATTCACCAAAAGAGGACATCACTGCAGA 2338
Qy 621 AlaCysGlyLysValPheCysAlaSerCysCysSerLeuLysCysLysLeuLysMet 640
Db 2339 GCATGTGGAAAGGTTTCTGTGCTTCGCTGAGGCTGAATGTAATCTTTATCATG 2398
Qy 641 AspArgLysGluAlaArgValCysValIleCysHisSerValLeuMetAsnAlaGlnAla 660
Db 2399 GACAGAAAGGAGCTAGAGTGTGTAACTCTGCCATTCACTGCTAAATGAAT 2449
Qy 661 TrpGluAsnMetSerAlaSerSerGlnSerProAsnProAsnAsnProAlaGluTyr 680
Db 2449 2449
Qy 681 CysSerThrIleProProLeuGlnGlnAlaGlnAlaSerGlyAlaLeuSerSerProPro 700
Db 2449 2449
Qy 701 ProThrValMetValProValGlyValLeuLysHisProGlyAlaGluValAlaGlnPro 720
Db 2450 2450
Qy 721 ArgGluGlnArgValTrpPheAlaAspGlyIleLeuProAsnGlyGluValAlaAsp 740
Db 2462 AGAGACGACAGCGAGTTGGTTTGTGATGGATCTTGGCCAAATGAGAGTTGCTGAT 2521
Qy 741 AlaAlaLysLeuThrMetAsnGlyThrSerSerAlaGlyThrLeuAlaValSerHisAsp 760
Db 2522 GCAGCCAAATTAACAATGAATGAATCTCTCTGAGGAACCTCTGGGTGTGTACACGAC 2581
Qy 761 ProValLysProValThrThrSerProLeuProAlaGluThrAspIleCysLeuPheSer 780
Db 2582 CCAGTCAAGCAGTAACCTACAGTCTCTACAGCAGAGCGATATTTGCTATTCTCT 2641
Qy 781 GlySerIleThrGlnValGlySerProValGlySerAlaMetAsnLeuIleProGluAsp 800
Db 2642 GGGAGTATAACTCAGTTGGAGTCTCTGTGAAGTGCATGAATCTTATTCTCTGAAGAT 2701
Qy 801 GlyLeuProProIleLeuIleSerThrGlyValLysGlyAspTyrAlaValGluGluLys 820
Db 2702 GGCCTTCCTCCCATTTCTCATCTCCATGGTGTAAAGAGAGACTATGCTGTGAAGAGAAA 2761
Qy 821 ProSerGlnIleSerValMetGlnGlnLeuGluAspGlyGlyProAspProLeuValPhe 840
Db 2762 CCATCAGAGATTTCAATGATGCAGCAGTTGGAGGATGGTGGCCCTGACCCACTTGTATT 2821
Qy 841 ValLeuAsnAlaAsnLeuLeuSerMetValLysIleValAsnTyrValAsnArgLysCys 860
Db 2822 GTTTTAAATGCAAAATTTCTGTCAATGGTTAAATTTGTAATTTATGTGAACAGGAAGTGC 2881
Qy 861 TrpCysPheThrThrLysGlyMetHisAlaValGlyGlnSerGluIleValIleLeuLeu 880
Db 2882 TGGTGTTCACAAACCAAGGAATGCATGCAGTGGTCACTCTGAGATAGTCACTTCTCTA 2941
Qy 881 GlnCysLeuProAspGluLysCysLeuProLysAspIlePheAsnHisPheValGlnLeu 900
Db 2942 CAGTGTTTACCGGATGAAAGTGTTCGCAAGGATATCTTTAAATCACTTTGTGCGAGCTT 3001
Qy 901 TyrArgAspAlaLeuAlaGlyAsnValValSerAsnLeuGlyHisSerPhePheSerGln 920
Db 3002 TATCGGGATGCTCTGGCAGGAATGTGTGAGCAACTTGGGACATTCCTTTCTTCAGTCAA 3061
Qy 921 SerPheLeuGlySerLysGluHisGlyGlyPheLeuTyrValThrSerThrTyrGlnSer 940
Db 3062 AGTTTCTTGGCAGTAAGAACATGGTGTGATTCCTATATGTGACATCTACCTACCACTCA 3121
Qy 941 LeuGlnAspLeuValLeuProThrProTyrLeuPheGlyIleLeuIleGlnLysTrp 960
Db 3122 CTGCAAGACCTAGTACTCCCAACCCCACTTACTTGTGTGGATTCCTTATCCAGAAATGG 3181

US-09-744-167-2 (1-1323) x AF104305 (1-4823)

Qy 1 MetTrpIleAspGluAsnAlaValAlaGluAspGlnLeuIleLysArgAsnTyrSerTrp 20
Db : : : : :
325 ATGTGTATGAGACACCATATATGCGAGAGATTCACTTTTAAACAACACTGGTCCCTCT 384
Qy 21 AspAspGlnCysSerAlaValGluValGluGluLysCysGlyAsnLeuAlaCysLeu 40
Db : : : : :
385 GAAATTGTATCAACCCATCTACTGAGAGTCAA-----AGTTTCAAGCTTTA 435
Qy 41 ProAspGluLysAsnValLeuValAlaValMethHisAsnCysAspLysArgThrLeu 60
Db : : : : :
436 GATGATCAATCA-----GTGAATATTCACATGAAAAAGTGTCTCTC 480
Qy 61 GlnAsnAspLeuGlnAspCysAsnAsnTyrAsnSerGlnSerLeuMetAspAlaPheSer 80
Db : : : : :
481 GCTGATGGCTTTTCCACCATGCAGT-----AGCCCCAAAAGTATTATAAAC----TTTGAC 531
Qy 81 Cys---SerLeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGlu 99
Db : : : : :
532 TGCCTGACCATGATACGAAATGCCCTTGCACAGTCAATGAGTGTGTGATGACATGAC 591
Qy 100 SerThrGluLysAspMetAsnSerGluLysGlnMetAspProLeuAsnArgProLysThr 119
Db : : : : :
591 : : : : :
Qy 120 GluGlyArgSerValAsnHisLeuCysProThrSerSerAspSerLeuAlaSerValCys 139
Db : : : : :
592 : : : : :
Qy 140 SerProSerGlnLeuLysAspAspGlySerIleGlyArgAspProSerMetSer----- 157
Db : : : : :
616 CTTCACAACATCATACAGGATACAGTAAGTAAAGCAGCAGCAGCAGTATCAATAAACCT 675
Qy 158 : : : : :
Db : : : : :
676 GGCCTAAAGAACCCCATAGAGCATTAAGAAACACCATCACTATTTCGCCTGAAATA 735
Qy 171 GlnGlyThrAspGlyCysProAlaValLys---LysGlnGluAsnTyrIleProAspGlu 189
Db : : : : :
736 AAGCCTTACTCCACATGTGCTGCCCTTCGTTTGAAATAACATAAGGTATCCACATTAT 795
Qy 190 AspLeuThrGlyLysIleSerSerProArgThrAspLeuGlySerProAsnSerPheSer 209
Db : : : : :
796 CAATTAAAT-----AATACAGATCTA-----CTCAGCGTTTCA 828
Qy 210 HisMetSerGluGlyIleLeuMetLysLysGluProAlaGluGluSerThrThrGluGlu 229
Db : : : : :
829 CCAGTGTGTTGAAGCATGTAGTACAGCAGCAGCAAAACATACATCTCTCTTGCATGAAGAA 888
Qy 230 SerLeuArgSerGlyLeuProLeuLeuLysProAspMetProAsnGlySerGlyArg 249
Db : : : : :
889 AAACCTTTTGAAGGTGT-----TCTGCAACG 915
Qy 250 AsnAsnAspCysGluArgCysSerAspCysLeuValProAsnGluValArgAlaAspGlu 269
Db : : : : :
916 GAGTCTTTGAGCCACTGTGCGGAAACTGTACTGGATAAT----- 957
Qy 270 AsnGluGlyTyrGluHisGluGluThrLeuGlyThrThrGluPheLeuAsn---MetThr 288
Db : : : : :
958 : : : : :
Qy 289 GluHisPheSerGluSerGlnAspMetThrAsnTrpLysLeuThrLysLeuAsnGluMet 308
Db : : : : :
997 AAGAACTTTTCTGACTCT---TGTGTATTAATGCGGACTTGACTAAAAGTTGTGCGCTC 1053
Qy 309 AsnAspSerGlnValAsnGluGluLysPheLeuGlnIleSerGlnProGluAsp 328
Db : : : : :
1054 TCT-----CAAGAAAGCAATGAAAGTTTGTGCAAGTAAAGATTGAA--- 1098
Qy 329 ThrAsnGlyAspSerGlyGlyGlnCysValGlyLeuAlaAspAlaGlyLeuAspLeuLys 348
Db : : : : :
1099 : : : : :
Qy 349 GlyThrCysIleSerGluSerGluCysAspPheSerThrValIleAspThrProAla 368
Db : : : : :
1129 AATGCTATGTAGCTTATAAAGAACAATAGATTGCTGAAGAAAATGGAACTAATGCA 1188
Qy 369 AlaAsnTyrLeuSerAsnGlyCysAspSerTyrGlyMetGlnAspProGlyValSerPhe 388
Db : : : : :
1189 CCAATGCTCTGTACAAATGGTGTGATTCCTATGGAATGAAAAACCCAGCCGTAGCTCAA 1248
Qy 389 ValProLysThrLeuProSerLysGluAspSerValThrGluGluLysGluIleGluGlu 408
Db : : : : :
1249 AACCCAAAAGAAATTTACCTTCAAAAGAAATCTGTGTGACAGAGAAAAAGAAATTTGAAGAA 1308
Qy 409 SerLysSerGluCysTyrSerAsnIleTyrGlu---GlnArgGlyAsnGluAlaThrGlu 427
Db : : : : :
1309 AGCAAGTCAGAAATCATCTACTGGTGTATTAACAACAAGAGAGATGATTACAGAG 1368
Qy 428 GlySerGlyLeuLeuLeuAsnSerThrGlyAspLeuMetLysLysAsnTyrLeuHisAsn 447
Db : : : : :
1369 AGAGGTGGACTTCTCTTAAATGCTAAGGCTCACCAATGAAGAAACAAT---TTGCATAGT 1425
Qy 448 PheCysSerGlnValProSerValLeuGlyGlnSerSerProLysValValAlaSerLeu 467
Db : : : : :
1426 CTTTGTATATCAGGTTCCATCCATGGGCAACATCACCAAAAAAGGGCAAGATTGTG 1485
Qy 468 ProSerIleSerValProPheGlyValAlaArgProLysGlnProSerAsnLeuLysLeu 487
Db : : : : :
1486 CAATCTCTCAGTGTTCATACGGTGGAGCACGCACTAAGCAGCCCACTCATCTCAAACCTC 1545
Qy 488 GlnIleProLysProLeuSerAspHisLeuGlnAsnAsp---PheProAlaAsnSerGly 506
Db : : : : :
1546 CATATTCCAAAGCCATTCTCTGAATGTGTGAGAGCGATCTCATCTCTCCAAATGCTGGC 1605
Qy 507 AsnAsnThrLysAsnLysAsnAspIleLeuGlyLysAlaLysLeuGlyGluAsnSerAla 526
Db : : : : :
1606 TGCAGCTCTTAATACAAAAAAGATGTTAAACAATCAATCAGGGGGAT----- 1656
Qy 527 ThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal-----Asp 541
Db : : : : :
1657 ---AACCTGATTTCAGAATCACTGCTGAGGATCTGAGTGGCAGCCCTGTACTCATGAT 1713
Qy 542 ThrAsnGlyGluHisLeuGluSerTyrGluAlaGluIleSerThrArgProCysLeuAla 561
Db : : : : :
1714 GCTAATGTGTATTTCCCTGGAGAAATACAGGGGACCTGGCAGCTTG-----TGCTTTGCA 1767
Qy 562 LeuAlaProAspSerProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAlaArg 581
Db : : : : :
1768 GTGTCTCCAGACAGCCAGACACAGATCTGCTTCCGGGGAGTTTGGGTACCCCATCTCT 1827
Qy 582 LysProPheThrThrLeuGlyGluValAlaProValTrpValProAspSerGluAlaPro 601
Db : : : : :
1828 AAGCCATTACTCTACTAGGGGAAGTGGCTCCAGTCTGGGTGGCAGATTCCCAAGCACC 1887
Qy 602 AsnCysMetLysCysGluAlaArgPheThrLysArgArgHisHisCysArgAla 621
Db : : : : :
1888 AACTGCATGAAGTCGAGGCCAGATTACATTATCCAAAAGAGGCGCATCACTGCGAGCT 1947
Qy 622 CysGlyLysValPheCysAlaSerCysCysSerLeuLysCysLysLeuLysMetAsp 641
Db : : : : :
1948 TGTGAAAAGGTGTTCTGTGCTGCTTGTGTCAGTCTAAATGCAAACTACAGTACATGGAT 2007
Qy 642 ArgLysGluAlaArgValCysValIleCysHisSerValLeuMetAsnAlaGlnAlaTrp 661
Db : : : : :
2008 AAAAAGGAGGCTCGTGTGTGTGTATTGTCTGCTTATGATGCTCAAGCATGG 2067
Qy 662 GluAsnMetMetSerAlaSerGlnSerProAsnProAsnProAlaGluTyrCys 681
Db : : : : :
2068 GAGAACATGTTAAGTGCATCGTCCAAAGCCCAATCAAAATAATCTCTGCTGAATACTGC 2127
Qy 682 SerThrIleProProLeuGlnAlaGlnAlaSerGlyAlaLeuSerSerProProPro 701
Db : : : : :
2128 TCACTATCTCCATCGAGGACACAGCTTCAGAGCACTGAGTTCCCACTCC 2187

QY	702	ThrValMetValProValGlyValLeuLysHisProGlyValAlaGluValAlaGlnProArg	721
DB	2188	ACTGTCATGGTGCAGTGGGTGCTTAAACAATCCAGGAACGAAGGGTCACAGTCAAG	2247
QY	722	GluGlnArgArgValTrpPheAlaAspGlyIleLeuProAsnGlyGluValAlaAspAla	741
DB	2248	GAACAGCGCGTGTGGTGTGCTGATGGAATATTACCAACGAGAGAGCTGCTGACTCA	2307
QY	742	AlaLysLeuThrMetAsnGlyThrSerSerAlaGlyThrLeuAlaValSerHisAspPro	761
DB	2308	GAT-----AATGCAACAGTAATACAGTGGCTGGGACACTTACTGTGTGCACAT	2355
QY	762	ValLysProValThrThrSerProLeuProAlaGluThrAsp---IleCysLeuPheSer	780
DB	2356	-----ACCAACAAATCCACATCTTCAGAGTCTGAGAACACCTCTGGATCTGT	2403
QY	781	GlySerIleThrGlnValGlySerProValGlySerAlaMetAsnLeuIleProGluAsp	800
DB	2404	GGAAGTATACTCAG-----GTTGGCAGTGCATAGAACCTTATTTCAGAAAGAT	2451
QY	801	GlyLeuProProIleLeuIleSerThrGlyValLysGlyAspTrpAlaValGluGluLys	820
DB	2452	GGGCTTCCTCCTATCTAATCTCTACTGGASTAAAGGAGATTACGCAGTTGAGGAACG	2511
QY	821	ProSerGlnIleSerValMetGlnGlnLeuGluAspGlyGlyProAspProLeuValPhe	840
DB	2512	CCITCCAGATGCTGTGATGCAGCACTAGAGGAGAGGAGCAGATCCTTTGGTTT	2571
QY	841	ValLeuAsnAlaAsnLeuLeuSerMetValLysIleValAsnTrpValAsnArgLysCys	860
DB	2572	GTTCTAAATGCAATCTTTGGCCATGGTTAAGATCGTGAACATGTTTACAGGAATGC	2631
QY	861	TrpCysPheThrThrLysGlyMetHisAlaValGlyGlnSerGluIleValIleLeuLeu	880
DB	2632	TGGTGCTTTACTACAAAGGAATGCATGCAGTGGCCAGGCTGAGATCGTAATCCTGTTG	2691
QY	881	GlnCysLeuProAspGluLysCysLeuProLysAspIlePheAsnHisPheValGlnLeu	900
DB	2692	CAGTGCCTGCCTGATGAGAAAGTGCTCCGAGGAGCACTGTTTAGCCATTTGTGTAGCTG	2751
QY	901	TyrArgAspAlaLeuAlaGlyAsnValValSerAsnLeuGlyHisSerPhePheSerGln	920
DB	2752	TATCAGAGGCAATATGCAGGTAAATGTAGTGGGAACTTGGGGCATTCCTCTCAGCCAG	2811
QY	921	SerPheLeuGlySerLysGluHisGlyPheLeuTyrValThrSerThrTyrGlnSer	940
DB	2812	AGTTTCCTGGGTAGTAAGGATCATGGTGGATTCTTTATGTTGCACCAACTACCACTCC	2871
QY	941	LeuGlnAspLeuValLeuProThrProProTyrLeuPheGlyIleLeuIleGlnLysTrp	960
DB	2872	CTCCAGACCTGGTCTCTCTCGACAGCCGCTACTGTTTGGAACTCTTATTCAAAAGTGG	2931
QY	961	GluThrProTrpAlaLysValPheProIleArgLeuMetLeuArgLeuGlyAlaGluTyr	980
DB	2932	GAGACTCCATGGCGCAAAAGTGTCCCATTCGGCTTATGCTGCGTTTAGGTGCAGATAC	2991
QY	981	ArgLeuTyrProCysProLeuPheSerValArgPheArgLysProLeuPheGlyGluThr	1000
DB	2992	AGATTGACCCATGTCACCTCTTCAGTGTTCGATACAGAAACCTCTGTTGGGGAACCC	3051
QY	1001	GlyHisThrIleMetAsnLeuLeuAlaAspPheArgSerTyrGlnTyrThrLeuProVal	1020
DB	3052	GGACACACCATCAATTAATGTTCTTAGCCGATTTTCAGAACTATCATGATATCTCTGCCAGT	3111
QY	1021	ValGlnGlyLeuValValAspMetGluValArgLysThrSerIleLysIleProSerAsn	1040
DB	3112	GTGCAGGGCTTGGTGGTATGGAAGTCAGAAAAAATAACATATAAAATCCCCAGCAAT	3171
QY	1041	ArgTyrAsnGluMetMetLysAlaMetAsnLysSerAsnGluHisValLeuAlaGlyGly	1060
DB	3172	AGATACATATGAGATCATGAAGCAATGAACAAATCCATGACATGCTGTGGCCATAGGA	3231

Qy	1061	AlaCysPheAsnGluLysAlaLeuSerHisLeuValCysValGlnAsnAspAspGlyAsn	1080
Db	3232	GCATGCTCAACACAGATGGCAGACTCTCACTTGTGTGTGTCAAACACGATGATGCAAT	3291
Qy	1081	TyrGlnThrGlnAlaIleSerIleHisAsnGlnProArgLysValThrGlyAlaSerPhe	1100
Db	3292	TACCAGACCAGGCAATTAGTATCCAAACAACACGCTAAAGTGCAGCGGGCCAGCTTC	3351
Qy	1101	PheValPheSerGlyAlaLeuLysSerSerSerGlyTyrIleuAlaLysSerIleVal	1120
Db	3352	TTTGTCTTCAGTGGTGCACCTAAAGTCTTCTTCGGGATACCTGGCCAAATCCAGCATAGTA	3411
Qy	1121	GluAspGlyValMetValGlnIleThrAlaGluAsnMetAspSerIleuArgGlnAlaLeu	1140
Db	3412	GAAGATGGGGTAATGTTTCAGATCACCGCAGAGAGATGGATGCCCTTCACACAGTCCCTT	3471
Qy	1141	ArgGluMetLysAspPheThrIleThrCysGlyLysAlaAspAlaGluGluProGlnGlu	1160
Db	3472	CGGGAGATGNAGGATTTCCACATTACATGTGGAAAGCTGATGCAGAGGAGTCACAGGAA	3531
Qy	1161	HisIleHisIleGlnTrpValAspAspAspLysAsnValSerLysGlyValValSerPro	1180
Db	3532	CATGCTCATGTCAGTGGGTGGAGGATGACAGAACTTTAACAAAGAGAGTTTTTAGTCCA	3591
Qy	1181	IleAspGlyLysSerMetGluThrIleThrAsnValLysIlePheHisGlySerGluTyr	1200
Db	3592	ATCGATGGCAATCAATGGAGTCTGTGCACCGCGTCAAGATTTTTCATGCTCAGAATAC	3651
Qy	1201	LysAlaAsnGlyLysValIleArgTrpThrGluValPhePheLeuGluAsnAspAspGln	1220
Db	3652	AAAGCTAGTGGAAAATAATTCGCTGGATAGAGTCTTCTTCTGGACAATGAGGAGCAA	3711
Qy	1221	HisAsnCysLeuSerAspProAlaAspHisSerArgLeuThrGluHisValAlaLysAla	1240
Db	3712	CAGAGTGGCCTGAGTGACCTGCTGATCATCAGCCGACTCACTGAAATGTGGCCAAAGCA	3771
Qy	1241	PheCysLeuAlaLeuCysThrGlnLeuLysLeuLeuLysGlyAspGlyMetThrLysLeu	1260
Db	3772	TTCTGTTTACGGCTTGGCCACACCTCAAGCTACTGAAGGAAGATGGAATGACCAAGTTA	3831
Qy	1261	GlyLeuArgValThrLeuAspSerAspGlnValGlyTyrGlnAlaGlySerAsnGlyGln	1280
Db	3832	GGTCTGGGGTGCATCTGGACTCAGACCAAGTGTGGATACCAAGCTGGGAGCAATGGGCAA	3891
Qy	1281	HisLeuProSerGlnTyrMetAsnAspPheAspSerAspLeuValLysMetIleHisGly	1300
Db	3892	CTCCTGCTGCCCGATACACCAATGATTTGGATGGTGCTTTGGTACCAAGTGATACACGGG	3951
Qy	1301	GlyAlaCysGlnLeuSerGluGlyProValValMetGluLeuIlePheTyrIleLeuGlu	1320
Db	3952	GGCACATGCCAGTTAAGTGAAGGCCCTGTTCAGTATGGAGCTGATATTTTATATCTCTTGAG	4011
Qy	1321	AsnIle	1322
Db	4012	AACATC	4017

RESULT 5	AF130420	2625 bp	linear	PRI 23-MAY-1999
LOCUS	AF130420		mrna	
DEFINITION	Homo sapiens serine protease-like protein isoform (NSP) mRNA, alternatively spliced, complete cds.			
ACCESION	AF130420			
VERSION	AF130420.1	GI:4883897		
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 2625)			
AUTHOR	Meckelein, B., Marshall, D.C., Conn, K.J., Pietropaolo, M., Van Norder, W., and Abraham, C.R.			
TITLE	Identification of a novel serine protease-like molecule in human brain			


```
Db 1799 AAGAAAAATATTATACATAATTTCTGTAGTCAAGTTCATCAGTGTGGCAATCTCC 1858
Qy 461 ProLysValValAlaSerLeuProSerIleSerValProPheGlyGlyAlaAaGProLys 480
Db 1859 CCCAGGTAGTAGCAAGCCCTGCCATCTATCAGTGTTCCTTTGGTGGTCAAGCCACAG 1918
Qy 481 GlnProSerAsnLeuLysLeuGlnIleProLysProLeuSerAspHisLeuGlnAsnAsp 500
Db 1919 CAACCTTCTAATCTTAACTTCAATTCAAATCCAAAGCCATATACAGACCATTTTACAAAATGAC 1978
Qy 501 PheProIleAsnSerGlyAsnAsnThrLysAsnLysAsnAspIleLeuGlyLysAlaLys 520
Db 1979 TTTCTCTCAACACAGTGGAAATATATCTAAATAATAAATAATATATCTTGGGAAGCAAA 2038
Qy 521 LeuGlyGluAsnSerAlaThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal 540
Db 2039 TTAGGGGAAACACTCAGCAACCAATGTATGAGTCCATCTTTGGGAAACATCTCTAATGTC 2098
Qy 541 AspThrAsnGlyGluHisLeuGluSerTyrGluAlaGluIleSerThrArgProCysLeu 560
Db 2099 GATACAAATGGGAACATTTAGAAAGTTATGAGGCTGAGATCTCCACTAGACCATGCCTT 2158
Qy 561 AlaLeuAlaProAspSerProAspAsnAspLeuArgAlaGlyGlnPheGlyLysSerAla 580
Db 2159 GCATTAGCTCCAGATACCCAGATAATGATCTCAGAGCTGGTCAGTTTGGAAATTTCTGCC 2218
Qy 581 ArgLysProPheThrThrLeuGlyGluValAlaProValTrpValProAspSerGlnAla 600
Db 2219 AGAAAGCCATTCACCACTCTGGGTGAGTGGCTCCAGTATGGGTACCGGATTTCTCAGGCT 2278
Qy 601 ProAsnCysMetLysCysGluAlaArgPheThrLysArgArgHisCysArg 620
Db 2279 CCAATTTGCATGAATGTGAAGCAGGTTTACATTCACCAAGAGGAGGATCTAGCAGA 2338
Qy 621 AlaCysGlyLysValPheCysAlaSerCysCysSerLeuLysCysLysLeuLeuTyrMet 640
Db 2339 GCATGTGGGAAGGTTTCTGTGCTCTCTGCTGCTAGCCTGAAATGTAAACTGTTATACATG 2398
Qy 641 AspArgLysGluAlaArgValCysValIleCysHisSerValLeuMetAsnAlaGlnAla 660
Db 2399 GACAGAAGGAAGCTAGAGTGTGTGTAATCTGCCATTCAGTGCTAATGAATGTGAATAT 2458
Qy 661 TrpGluAsnMetMetSer 666
Db 2459 TAAACAGGTGTGATTCA 2476

RESULT 6
AX809313 1977 bp DNA linear PAT 25-NOV-2003
LOCUS Sequence 10 from Patent WO03045990.
DEFINITION AX809313
ACCESSION AX809313
VERSION AX809313.1 GI:38523628
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Legrain, P., Gauthier, J.M., Colland, F. and Jacq, X.
TITLE Protein-protein interactions involving transforming growth factor
beta signalling
JOURNAL Patent: WO 03045990-A 10 05-JUN-2003;
Hybridgenics (FR)
FEATURES
source
Location/Qualifiers
1..1977
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores: 1.1e-216 Length: 1977
Pred. No.:
```

```
Score: 3397.00 Matches: 650
Percent Similarity: 98.94% Conservative: 1
Best Local Similarity: 98.78% Mismatches: 7
Query Match: 48.60% Indels: 0
DB: 6 Gaps: 0
US-09-744-167-2 (1-1323) x AX809313 (1-1977)
Qy 666 SerAlaSerSerGlnSerProAsnProAsnAsnProAlaGluTyrCysSerThrIlePro 685
Db 1 AGTGCTCAAGCCAGAGCCCTAACCTAACATCTCTGCTGAATCTGTCTACTATCCCT 60
Qy 686 ProLeuGlnGlnAlaGlnAlaSerGlyValAlaLeuSerSerProProThrValMetVal 705
Db 61 CCCCTTGCAGCAAGCTCAGCCCTCAGAGCTCTGAGCTCTCCACTCCCACTGTGTGTGTA 120
Qy 706 ProValGlyValLeuLysHisProGlyAlaGluValAlaGlnProArgGluGlnArg 725
Db 121 CCTGTGGAGATTTTAAAGCACCCCTGGAGCAGAGTGGCTCAGCCAGAGAGCAGAGGGA 180
Qy 726 ValTrpPheAlaAspGlyLysLeuProAsnGlyGluValAlaAspAlaAlaLysLeuThr 745
Db 181 GTTGTGTGTGTGATGGGATCTTGCCCAATGGAGAAGTGTGTGATGCGAGCCAAATTAACA 240
Qy 746 MetAsnGlyThrSerSerAlaGlyThrLeuAlaValSerHisAspProValLysProVal 765
Db 241 ATGAATGGAACTTCTCTGCAGGAACCTGGCTGTGTGCACAGCCAGTCAAGCCAGTA 300
Qy 766 ThrThrSerProLeuProAlaGluThrAspIleCysLeuPheSerGlySerIleThrGln 785
Db 301 ACTACAGTCTCTCTACCAGCAGAGCGGATATTGTCTATTCTCTGGGAGATATAACTCAG 360
Qy 786 ValGlySerProValGlyLysAlaMetAsnLeuIleProGluAspGlyLeuProPhe 805
Db 361 GTTGGAACTCCTGTGGAAAGTGCATGAATCTTATCTCTGAGATGGCTTCTCTCCATT 420
Qy 806 LeuIleSerThrGlyValLysGlyAspTyrAlaValGluGluLysProSerGlnIleSer 825
Db 421 CTCTCTCCACTGGTGTAAAGAGACATATGTGTGGAAGAGAAACCATCACAGATTCA 480
Qy 826 ValMetGlnGlnLeuGluAspGlyGlyProAspProLeuValPheValLeuAsnAlaAsn 845
Db 481 GTAATGCAGCAGTGGAGATGGTGGCCCTGACCCACTGTATTTGTTTTAAATGCATAT 540
Qy 846 LeuLeuSerMetValLysIleValAsnTyrValAsnArgLysCysTyrCysPheThrThr 865
Db 541 TTGTTGTCAATGTTTAAATTTATAATTTATGTGAACAGAGAGTGTGGTGTTCACACC 600
Qy 866 LysGlyMetHisAlaValGlyGlnSerGluIleValIleLeuLeuGlnCysLeuProAsp 885
Db 601 AAGGGAATGCATGCGATGGGTGAGTCTGAGATAGTCAATCTTCTACAGTGTTCCTGGAT 660
Qy 886 GluLysCysLeuProLysAspIlePheAsnHisPheValGlnLeuTyrArgAspAlaLeu 905
Db 661 GAAAGTGTTCGCAAGAGATATCTTTAATCACTTTGTGACAGTTTATCGGATGCTCTG 720
Qy 906 AlaGlyAsnValLysSerAsnLeuGlyHisSerPhePheSerGlnSerPheLeuGlySer 925
Db 721 GCAGGAATGTGTGAGCAACTTGGGACATCTCTTCTTCAGTCAAAAGTTTCTCTGGCAGT 780
Qy 926 LysGluHisGlyGlyPheLeuTyrValThrSerThrTyrGlnSerLeuGlnAspLeuVal 945
Db 781 AAAGAACATGGTGGATTTCTTATGTGACATCTTACTACCACTACTGCAAGACCTAGTA 840
Qy 946 LeuProThrProProTyrLeuPheGlyIleLeuIleGlnLysTyrPheGluThrProTrpAla 965
Db 841 CTCCTCAACCCACCTTACTTTGTTGGATTTCTTATCCAGAAATGGAAATCTCTCTGGCT 900
Qy 966 LysValPheProIleArgLeuMetLeuArgLeuGlyAlaGluTyrArgLeuTyrProCys 985
Db 901 AAAGTATTTCCCTATCCGTCGTGATGTGAGACTTGGAGCTGAATATGATCTTATCATC 960
Qy 986 ProLeuPheSerValArgPheArgLysProLeuPheGlyGluThrGlyHisThrIleMet 1005
```


Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

EcoRI				BglII				HindIII			
SeqDerMap	EngPrnt	SeqDerMap	EngPrnt	SeqDerMap	EngPrnt	SeqDerMap	EngPrnt	SeqDerMap	EngPrnt	SeqDerMap	EngPrnt
5810	5855	7542	7542	1289	1229						
2184	2142	5671	5692	449	<800						
8065	8190	5077	5096	512	<800						
486	<800	76	<800	2814	2825						
106	<800	3560	3636	1247	1229						
4616	4708	414	<800	4695	4762						
14239	14077	1723	1693	3373	3356						
3925	3921	1216	1202	3274	3356						
983	984	2887	2916	3116	3125						
2594	2609	8059	7992	68	<800						
2871	2878	1854	1935	1886	1926						
2276	2254	4987	4925	3144	3125						
2711	2609	13600	12991	9189	9097						
4738	4708	2718	2784	1213	1229						
445	<800	2884	2916	1915	1926						
1531	1495	4547	4472	731	747						
3546	3538	467	<800	3148	3125						
3154	3144	900	919	3481	3356						
896	984	2356	2329	2570	2594						
802	884	134	<800	5185	5250						
11042	10836	4510	4472	868	865						
4907	4887	3070	3044	18109	18317						
6450	6524	2284	2329	12807	12493						
3079	3045	3751	3739	4972	4989						
532	<800	2586	2640	4206	4097						
773	777	1052	1063	374	<800						
1464	1495	1058	1063	1931	1926						
6249	6230	236	<800	11410	11255						

1821	1816	1089	1063	9089	9097
2545	2609	3536	3471	5486	5606
1540	1495	440	<800	92	<800
2491	2609	9271	9390	14507	14568
1578	1495	1482	1491	5171	5250
1510	1495	916	919	1939	1926
2118	2142	1271	1267	166	<800
1084	1065	340	<800	6161	6265
1697	1677	3217	3260		
3818	3823	2323	2329		
1156	1132	9366	9390		
10017	9930	159	<800		
4212	4165	3486	3471		
2425	2405	4158	4097		
2682	2609	1524	1491		
1017	984	9505	9390		
7981	8190	6646	6734		
421	<800	2639	2640		
FEATURES					
Location/Qualifiers					
1..133799					
/organism="Homo sapiens"					
/mol_type="genomic DNA"					
/db_xref="taxon:9606"					
/chromosome="1"					
/clone="RP4-814E15"					
/clone_lib="RPC1 human PAC library 4"					
10029..10030					
/note="IS1 and 9-bp target site duplication excised from between bases 10029 and 10806"					
GGTAATGACTCCAACTTACTGATAGTGTTTATGTCAGATAATCCCGATGACCTTG					
TCATGCAACTCCACCGATTTCAGAACACAGACTTCCGTCCAGCTTCCAGAT					
GTGTCTCAGATTTCAGATTATCGCTCAATCGCTGAGTGAACGCTTGCTGATAAC					
misc_feature					
Alignment Scores:					
Pred. No.: 8.4e-210 Length: 133799					
Score: 3333.00 Matches: 632					
Percent Similarity: 96.95% Conservative: 3					
Best Local Similarity: 96.49% Mismatches: 8					
Query Match: 47.68% Indels: 12					
DB: Gaps: 2					
US-09-744-167-2 (1-1323) x AC105754 (1-133799)					
QY 1 MetTTPileAspGluAsnAlaValAlaGluAspGlnLeuLeuLysArgAsnTyrSerTTP 20					
Db 60325 ATGTGATTGATGAAATGCTGTTGCAGAACCACTTAATTAAGAGAACTATAGTTGG 60384					
QY 21 AspAspGlnCysSerAlaValGluValGlyGluLysCysGlyAsnLeuAlaCysLeu 40					
Db 60385 GATGATCAATGACGTCTGTTGAAGTGGAGAGAGAAATGGAAACCTGGCTTGTCTG 60444					
QY 41 ProAspGluLysAsnValLeuValAlaValMetHisAsnCysAspLysArgThrLeu 60					
Db 60445 CCAGATGAGAGAGATGTTCTTGTGTAGCCGTCATGCAATGCTATTAAGGACATTA 60504					

QY 61 GlnAsnAspLeuGlnAspCysAsnAsnTyrAsnSerGlnSerLeuMetAspAlaPheSer 80
 Db 60505 CAAACAGATTATACAGATTGTAATTAATTAATAGTCAATCCCTATGCGTCTTTAGC 60564
 QY 81 CysSerLeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGluSer 100
 Db 60565 TGTTCACTGGTAATCAAAACAGACAAACTGATCAATTTAGTTAGTATATAATCACTCC 60524
 QY 101 ThrGluLysAspMetAsnSerGluLysGlnMetAspProLeuAsnArgProLysThrGlu 120
 Db 60625 ACTGAAGAAGATATGAATTCAGAGAAACAATGGATCCATTGAATAGACCCGAAACAGAG 60684
 QY 121 GlyArgSerValAsnHisLeuCysProThrSerSerAspSerLeuAlaSerValCysSer 140
 Db 60685 GGGAGATCTGTAAACCATCTGTCTACTTCACTCATCTAGATCTAGCCAGTGTCTGTTC 60744
 QY 141 ProSerGlnLeuLysAspAspGlySerIleGlyArgAspProSerMetSerAlaIleThr 160
 Db 60745 CTTTCACCAATTAAGGATGACGGAAGTATAGTAGACCCCTCCATGCTCTGCGATTACA 60804
 QY 161 SerLeuThrValAspSerValIleSerSerGlnGlyThrAspGlyCysProAlaValLys 180
 Db 60805 AGTTTAACGGTTGATTCAGTAATCTCATCCAGGGAACAGATGATGCTCTGCTGTAAA 60864
 QY 181 LysGlnGluAsnTyrIleProAspGluAspLeuThrGlyLysIleSerSerProArgThr 200
 Db 60865 AAGCAAGAGAATAATATACAGATGAGGACCTCACTCGCAAAATCAGCTCTCTCAGGACA 60924
 QY 201 AspLeuGlySerProAsnSerPheSerHisMetSerGluGlyIleLeuMetLysLysGlu 220
 Db 60925 GATCTAGGGAGTCCAAATTCCTTTTCCACATGATGAGGGGATTTTGATGAAACAGAG 60984
 QY 221 ProAlaGluGluSerThrThrGluSerLeuArgSerGlyLeuProLeuLeuLys 240
 Db 60985 CCAGCAGAGGAGAGCACCACCTGAAGAATCCCTCCGGTCTGGTTACCTTTGCTTCTCAA 61044
 QY 241 ProAspMetProAsnGlySerGlyArgAsnAspCysGluArgCysSerAspCysLeu 260
 Db 61045 CCAGCATGCTTAATGGTCTGGAGGAATTAATGACTGTGAACGGTCTTCCAGATTGCTT 61104
 QY 261 ValProAsnGluValArgAlaAspGluAsnGluGlyTyrGluHisGluGluThrLeuGly 280
 Db 61105 GTGCTTAATGAATTAGGGCTGATGAAATGAAGGTTTATGAACATGAAGAACTCTTGCC 61164
 QY 281 ThrThrGluPheLeuAsnMetThrGluHisPheSerGluSerGlnAspMetThrAsnTrp 300
 Db 61165 ACTACAGAAATTCCTTAATATACAGACAGCATTTCTCGAATCTCAGGACATGACTAATGG 61224
 QY 301 LysLeuThrLysLeuAsnGluMetAsnAspSerGlnValAsnGluLysGluLysPhe 320
 Db 61225 AAGTTGACTAACTAATAGATGAATGAATGATAGCAAGTAAACGAGAAAGAAAGTTT 61284
 QY 321 LeuGlnIleSerGlnProGluAspThrAsnGlyAspSerGlyGlnCysValGlyLeu 340
 Db 61285 CTACAGATTAGTACGCTGAGGACACTAATGGTATAGTGGAGGACAGTGTGTGATTG 61344
 QY 341 AlaAspAlaGlyLeuAspLeuLysGlyThrCysIleSerGluSerGluGluCysAspPhe 360
 Db 61345 GCAGATCGAGTCTAGATTAAAGGAACTTGATAGTGAAGTGAAGATGATGATTTC 61404
 QY 361 SerThrValIleAspThrProAlaAsnTyrIleSerAsnGlyCysAspSerThrGly 380
 Db 61405 TCCACTGTTATAGACACAGCAGCAATATCTATCTAATGGTGTGATTCCTATGGA 61464
 QY 381 MetGlnAspProGlyValSerPheValProLysThrLeuProSerLysGluAspSerVal 400
 Db 61465 ATGCAAGACCCAGGTGTTCTTTGTTCCAAAGACTTTACCTCCAAAGAAAGATTGATGA 61524
 QY 401 ThrGluGlyLysGluIleGluGluSerLysSerGluCysThrSerAsnIleThrGluGln 420
 Db 61525 ACAGAAGAAAAAGAAATAGAGGAAAGCAAGTCAAGATGCTACTCAATATTATGAACAG 61584
 QY 421 ArgGlyAsnGluAlaThrGluGlySerGlyLeuLeuAsnSerThrGlyAspLeuMet 440

Db 61585 AGAGGAATAGAGGCCACAGAGGGAGTGACTACTTTTAAACAGCAGCTGGTACCTAATG 61644
 QY 441 LysLysAsnTyrIleuHisAsnPheCysSerGlnValProSerValLeuGlyGlnSerSer 460
 Db 61645 AGAAAAATTAATTTACATAAATTTCTGTAGTCAAGTTCCATCAGTCTGGGCAATCTTCC 61704
 QY 461 ProLysValValAlaSerLeuProSerIleSerValProPheGlyGlyValArgProLys 480
 Db 61705 CCCAAGGTAGTAGAGGCTGCCATCTATCATGTTCCTTTTGGTGGTGGCAAGCCCAAG 61764
 QY 481 GlnProSerAsnLeuLysLeuGlnIleProLysProLeuSerAspHisLeuGlnAsnAsp 500
 Db 61765 CAACCTTCTAATCTTAAACTTCAAATTCAAAGCCATTATCAGACCAITTTACAAATGAC 61824
 QY 501 PheProAlaAsnSerGlyAsnAsnThrLysAsnLysAsnAspIleLeuGlyLysAlaLys 520
 Db 61825 TTTCTTGCACAAACAGTGGAAATAATACTAAATAAAATGATATTTCTTGGGAAAGCAAAA 61884
 QY 521 LeuGlyGluAsnSerAlaThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal 540
 Db 61885 TTAGGGGAAACTCAGCAACCAATGTATGCGATCCATCTTTGGGAAACATCTCTAATGTC 61944
 QY 541 AspThrAsnGlyGluHisLeuGluSerTyrGluAlaGluIleSerThrArgProCysLeu 560
 Db 61945 GATACAAATGGGAAACATTTAGAAAGTTATGAGGCTGAGATCTCCACTAGACCATGCTT 62004
 QY 561 AlaLeuAlaProAspSerProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAla 580
 Db 62005 GCATTAGTCTCCAGATAGCCAGATATGATCTCAGAGCTGGTCACTTTGGATTCTTGCC 62064
 QY 581 ArgLysProPheThrThrLeuGlyGluValAlaProValTrpValProAspSerGlnAla 600
 Db 62065 AGAAAGCCATTCCACACTCTGGTGGAGTGGCTCCAGTATGGTACCGGATTTCTCAGCT 62124
 QY 601 ProAsnCysMetLysCysGluAlaArgPheThrPheThrLysArgArgHisCysArg 620
 Db 62125 CCAAAATTCATGAAATGTGAAGCCAGGTTTACATTCACAAAGAGGAGCATCTCGAGA 62184
 QY 621 AlaCysGlyLysValPheCysAlaSerCysCysSerLeuLysCysLysLeuLeuMet 640
 Db 62185 GCATCTCGGAAAGGTAAGTTGC-----ATGTATACA 62214
 QY 641 AspArgLysGluAlaArgValCysValIleCysHisSerValLeu 655
 Db 62215 CTCAGAAAT-----CGGGCATGCACATTTGTATGCTGAATTA 62253

RESULT 8

AC026909

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL


```

misc_feature /note="assembly_fragment"
47639..53052
misc_feature /note="assembly_fragment"
53153..58427
misc_feature /note="assembly_fragment"
58528..67461
misc_feature /note="assembly_fragment"
67562..79052
misc_feature /note="assembly_fragment"
79153..92165
misc_feature /note="assembly_fragment"
92266..110031
misc_feature /note="assembly_fragment"
clone_end:SP6
vector_side:left"
110132..129806
misc_feature /note="assembly_fragment"
129907..150129
misc_feature /note="assembly_fragment"

ORIGIN
Alignment Scores:
Pred. No.: 9,91e-210 Length: 150129
Score: 3333.00 Matches: 632
Percent Similarity: 96.95% Conservative: 3
Best Local Similarity: 96.49% Mismatches: 8
Query Match: 47.68% Indels: 12
DB: 2 Gaps: 2

US-09-744-167-2 (1-1323) x AC026909 (1-150129)
QY 1 MetTpIleAspGluAsnAlaValAlaGluAspGlnLeuIleLysArgAsnTyrSerTirp 20
DB 70313 ATGTGGATTGATGAATGCTGTTCGACAGACCCAGTTAATTAAGAGAAACTATAGTTGG 70372
QY 21 AspAspGlnCysSerAlaValGluValGlyGluLysCysGlyAsnLeuAlaCysLeu 40
DB 70373 GATGATCAATGCAGTGTCTTGAAGTGGGAGAGAAATGTGGAAACCTGGCTGTCTG 70432
QY 41 ProAspGluLysAsnValLeuValAlaValAlaValMethIleAsnCysAspLysArgThrLeu 60
DB 70433 CCAGATGAGAGAAATGTTCTTGTGTAGCCGTCATGATCAATCTTATGATGATGATGAT 70492
QY 61 GlnAsnAspLeuGlnAspCysAsnAsnTyrAsnSerGlnSerLeuMetAspAlaPheSer 80
DB 70493 CAAACCGATTACAGGATTGTAATTAATTAATAGTCAATCCCTTATGGATGCTTTAGC 70552
QY 81 CysSerLeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGluSer 100
DB 70553 TGTTCACTGGATATGAATGAACAGACAACTGATCAATTTAGTTTATAGTATAAATGATGTC 70612
QY 101 ThrGluLysAspMetAsnSerGluLysGlnMetAspProLeuAsnArgProLysThrGlu 120
DB 70613 ACTGAAGAAGATATGATTCAGAGAAACAAATGATCCATTCATAGACCCAAACAGAG 70672
QY 121 GlyArgSerValAsnHisLeuCysProThrSerSerAspSerLeuAlaSerValCysSer 140
DB 70673 GGGAGATCTGTAAACCATCTGTGTCTTACTCTATCTATCTATCTATCTATCTATCTATCT 70732
QY 141 ProSerGlnLeuLysAspAspGlySerIleGlyArgAspProSerMetSerAlaIleThr 160
DB 70733 CCTTCACAAATTAAGGATGACGGAAGTATAGTAGAGACCCCTCCCATGCTCGGATTA 70792
QY 161 SerLeuThrValAspSerValIleSerSerGlnGlyThrAspGlyCysProAlaValLys 180
DB 70793 AGTTTAACCGTTGATTGATTAATCTCATCCAGGAACAGATGGATGTCCTGCTGTATA 70852
QY 181 LysGlnGluAsnTyrIleProAspGluAspLeuThrGlyLysLeuSerSerProArgThr 200
DB 70853 AAGCAGAGAACTATATACAGATGAGGACCTCAGTGGCAAAATCAGCTCTCCTAGGACA 70912
QY 201 AspLeuGlySerProAsnSerPheSerHisMetSerGluGlyIleLeuMetLysLysGlu 220

```

```

DB 70913 GATCTAGGGAGTCCAAATTCCTTTTCCCATCATGAGTGGGGATTTTGATGAAAAAGAG 70972
QY 221 ProAlaGluGluSerThrThrGluGluSerLeuAArgSerGlyLeuProLeuLeuLys 240
DB 70973 CCAGCAGAGAGAGACCACTGAAGAAATCCCTCCGGTCTCGGTTTACCTTTGCTTCTCAA 71032
QY 241 ProAspMetProAsnGlySerGlyArgAsnAsnAspCysGluArgCysSerAspCysLeu 260
DB 71033 CCAGACATGCCTAATGGGTCTGCAAGGAATAATGACTGTGAACGGTGTTCAGATTGCCCT 71092
QY 261 ValProAsnGluValArgAlaAspGluAsnGluGlyTyrGluHisGluGluThrLeuGly 280
DB 71093 GTCCCTAATGAAGTTAGGGCTGATGAATGAAGGTATGAACATGAGAAACTCTTGCC 71152
QY 281 ThrThrGluPheLeuAsnMetThrGluHisPheSerGluSerGlnAspMetThrAsnTirp 300
DB 71153 ACTACAGAATTCCTTAATATGACAGAGCATTTCTCTGAATCTCAGACATGACTAATGG 71212
QY 301 LysLeuThrLysLeuAsnGluMetAsnAspSerGlnValAsnGluGluLysGlyLysPhe 320
DB 71213 AAGTTGACTAAACTAAATGAGATGAATGATAGCCAAAGTAAACGAGAAAGAAAGTTT 71272
QY 321 LeuGlnIleSerGlnProGluAspThrAsnGlyAspSerGlyGlyGlnCysValGlyLeu 340
DB 71273 CTACAGATTAGTACGCCCTGAGGACACTAATGTGTATGTGGAGACAGTGTGTGGATTG 71332
QY 341 AlaAspAlaGlyLeuAspLeuLysGlyThrCysIleSerGluSerGluGluCysAspPhe 360
DB 71333 GCAGATGCAGGTCTAGATTAAAGGAACTTGTCATTAGTAAAGTGAAGATGTGATTTC 71392
QY 361 SerThrValIleAspThrProAlaAlaAsnTyrLeuSerAsnGlyCysAspSerTyrGly 380
DB 71393 TCACCTGTTATAGACACACCAGCAGCAATATATCTATCTAATGTTGTGATTCTCTATGGA 71452
QY 381 MetGlnAspProGlyValSerPheValProLysThrLeuProSerLysGluAspSerVal 400
DB 71453 ATGCAAGACCCAGGTGTTCTTTCTTCCAAAGACTTTTACCTCCAAAGAAAGATTCACTA 71512
QY 401 ThrGluGluLysGluIleGluGluSerLysSerGluCysTyrSerAsnIleTyrGluGln 420
DB 71513 ACAGAGAAAGAAAGAAATAGAGAAAGCAAGTCAGAAATCTCTACTCAATATATTATGAACAG 71572
QY 421 ArgGlyAsnGluAlaThrGluGlySerGlyLeuLeuLeuAsnSerThrGlyAspLeuMet 440
DB 71573 AGAGGAATAGGCCACAGAGAGGAGTGCATCTTTTAAACAGCACTGGTGACCTAATG 71632
QY 441 LysLysAsnTyrLeuHisAsnPheCysSerGlnValProSerValLeuGlyGlnSerSer 460
DB 71633 AAGAAAAATTAATTTACATAAATTTCTGTAGTCAAGTTCCATCAGTGTCTGGGCAATCTTCC 71692
QY 461 ProLysValValAlaSerLeuProSerIleSerValProPheGlyGlyAlaArgProLys 480
DB 71693 CCCAAGGTAGTAGCAGGCTGCCATCTATCACTGTTCTTTTGGTGGTGCAGACCCCAAG 71752
QY 481 GlnProSerAsnLeuLysLeuGlnIleProLysProLeuSerAspHisLeuGlnAsnAsp 500
DB 71753 CAACCTCTTAATCTTAAACTTTCAAAATTCAAAAGCCATTTATCAGACCATTTACAAAATGAC 71812
QY 501 PheProAlaAsnSerGlyAsnAsnThrLysAsnLysAsnAspIleLeuGlyLysAlaLys 520
DB 71813 TTTCTCGCAACAGTGGAAATAAATACATAAAATAAAATGATATATCTTTGGGAAGCAAAA 71872
QY 521 LeuGlyGluAsnSerAlaThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal 540
DB 71873 TTAGGGGAAAACTCAGCAACCAATGTATGCACTCCATCTTTTGGGAAACATCTCTAATGTC 71932
QY 541 AspThrAsnGlyGluHisLeuGluSerTyrGluAlaGluIleSerThrArgProCysLeu 560
DB 71933 GATACAAATGGGGAACATTTTAGAAGTTATAGAGCTGAGATCTCCACTAGACCATGCTCT 71992
QY 561 AlaLeuAlaProAspSerProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAla 580
DB 71993 GCATTAGTCCAGATAGCCCAAGATAATGATCTCAGAGCTGGTCAGTTTGGAAATTTCTGCC 72052

```

QY 581 ArgLysProPheThrThrLeuGlyGluValAlaProValTrpValProAspSerGlnAla 600
 Db 72053 AGAAGCCATTACCACTCTGGGTAGGTGGTCCAGTATGGTACCGGATTCTCAGGCT 72112

QY 601 ProAsnCysMetLysCysGluAlaArgPheThrPheThrLysArgArgHisCysArg 620
 Db 72113 CCAAAATGGATGAATGTGAAGCCAGGTTTACATTACCAAAAGAGGAGCATCACTGCAGA 72172

QY 621 AlaCysGlyLysValPheCysAlaSerCysCysSerLeuLysCysLysLeuLeuTrMet 640
 Db 72173 GCATGTGGGAAGTAAGTTGC-----ATGTATACG 72202

QY 641 AspArgLysGluAlaArgValCysValIleCysHisSerValLeu 655
 Db 72203 CTCAGAAAT-----CGGCATGCACATTGTGATGTCGAATTA 72241

RESULT 9
 AL591398
 LOCUS Human DNA sequence from clone RP11-81C12 on chromosome X, complete sequence.
 DEFINITION AL591398
 ACCESSION AL591398.2 GI:14330165
 VERSION HTG.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 89973)
 AUTHORS Heath, P.
 TITLE Direct Submission
 JOURNAL Submitted (05-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 CLONE REQUESTS: clonerequests@sanger.ac.uk
 COMMENT On Jun 8, 2001 this sequence version replaced gi:1433279.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
 This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/ChrX>
 RP11-81C12 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6
 IMPORTANT: This sequence is not the entire insert of clone RP11-81C12 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
 The true left end of clone RP3-389B13 is at 89874 in this sequence.
 The true right end of clone RP1-137H15 is at 100 in this sequence.
 The true right end of clone RP1-7706 is at 44281 in this sequence.

FEATURES
 source Location/Qualifiers
 1..89973
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

```

repeat_region 13268..13327
/note="30 copies 2 mer ta 98% conserved"
repeat_region 14758..14984
/note="MIR repeat: matches 16..257 of consensus"
repeat_region 15103..15315
/note="12 repeat: matches 1551..1774 of consensus"
repeat_region 15324..15625
/note="12 repeat: matches 2442..2750 of consensus"
repeat_region 16017..16079
/note="MIR repeat: matches 78..139 of consensus"
repeat_region 16085..16215
/note="MER33 repeat: matches 704..1687 of consensus"
repeat_region 16594..16617
/note="12 copies 2 mer tg 95% conserved"
repeat_region 16923..16980
/note="29 copies 2 mer aa 74% conserved"
repeat_region 17294..17497
/note="MIR repeat: matches 12..254 of consensus"
misc_feature 17730..18153
/note="match: GSS: Em:AQ264531"
misc_feature 17743..17945
/note="match: GSS: Em:AQ263645"
repeat_region 18255..18278
/note="12 copies 2 mer gt 100% conserved"
repeat_region 18578..18872
/note="Aluub repeat: matches 1..294 of consensus"
repeat_region 19101..19464
/note="MER39 repeat: matches 13..388 of consensus"
repeat_region 19508..19598
/note="12 repeat: matches 2662..2750 of consensus"
repeat_region 20288..20583
/note="Aluub repeat: matches 1..293 of consensus"
repeat_region 21186..21309
/note="62 copies 2 mer tt 59% conserved"
repeat_region 21665..21708
/note="22 copies 2 mer ac 77% conserved"
repeat_region 23271..23419
/note="MIR repeat: matches 37..173 of consensus"
repeat_region 23420..23720
/note="Aluub repeat: matches 3..310 of consensus"
repeat_region 23721..23800
/note="MIR repeat: matches 173..255 of consensus"
repeat_region 23906..23960
/note="MLT2B repeat: matches 394..448 of consensus"
repeat_region 23961..24050
/note="45 copies 2 mer ta 88% conserved"
repeat_region 24182..24576
/note="MLT2B repeat: matches 1..404 of consensus"
repeat_region 25118..25401
/note="L1MB1 repeat: matches 5747..6063 of consensus"
misc_feature 28019..28534
/note="match: GSS: Em:A2519577"
repeat_region 28603..28695
/note="MIR repeat: matches 52..146 of consensus"
repeat_region 29276..29754
/note="12 repeat: matches 1891..2393 of consensus"
repeat_region 29933..30218
/note="Aluub repeat: matches 1..286 of consensus"
repeat_region 31138..31447
/note="AluSc repeat: matches 1..309 of consensus"
repeat_region 32120..32398
/note="MER7A repeat: matches 62..346 of consensus"
repeat_region 32401..32510
/note="55 copies 2 mer ta 72% conserved"
repeat_region 32533..32560
/note="14 copies 2 mer ac 100% conserved"
repeat_region 32561..32635
/note="78K repeat: matches 1..73 of consensus"
repeat_region 33629..33658
/note="15 copies 2 mer ct 90% conserved"
repeat_region 34128..34286
/note="MIR repeat: matches 95..262 of consensus"
repeat_region 34349..34684

```

```

Alignment Scores: 9.2e-163 Length: 89973
Prod. No.: 2623.00 Matches: 525
Score: 2623.00 Conservative: 11
Percent Similarity: 90.36% Mismatches: 38
Best Local Similarity: 90.36% Indels: 9
Query Match: 37.53% Gaps: 2
DB: 9
US-09-744-167-2 (1-1323) x AL591398 (1-89973)

QY 737 GluValAlaAspAlaAlaLysLeuThrMetAnGlyThrSerSerAlaGlyThrLeuAla 756
Db 51984 GAAGTGTCTGATGCAGCCAAATAGCAATGAATGAATCTCTCTGCAGGACCCCTGGCT 52043
QY 757 ValSerHisAspProValLysProValThrThrSerProLeuProAlaGluThrAspIle 776
Db 52044 GTGTACACGACCCAGTCAAGCTAGTACTACACAGTCTCTACACGACGAGATGGATATT 52103
QY 777 CysLeuPheSerGlySerIleThrGlnValGlySerProValGlySerAlaMetAsnLeu 796
Db 52104 CCTCTATTCTCTGGCAGTATTAATCAGGTTCAGAGATCTCTGTGGAGTGCATGAGCCTT 52163
QY 797 IleProGluAspGlyLeuProProIleLeuIleSerThrGlyValLysGlyAspIleAla 816
Db 52164 ATTCTGAAGATGGCCTTCTCTCTATTCTCATCTCTCACTGGTGTAAAGGAGACTATGCT 52223
QY 817 ValGluGluLysProSerGln----IleSerValMetGlnGlnLeuGluAspGlyGlyPr 835
Db 52224 GTGGAAGATATAACCATCACAGATTTATTTCAGAAATGCAGCAGTTGGAAGACGCTGGCCC 52283
QY 835 OAspProLeuValPheValLeuAsnAlaAsnLeuSerMetValLysIleValAsnTy 855
Db 52284 TGACCCACCTGTATTGTTTAAATGCAAAATTTGTGCAATGGTTAAATTTGAAATTA 52343
QY 855 rValAsnArgLysCysTyrCysPheThrThrLysGlyMethHisAlaValAlGlyGlnSerGl 875
Db 52344 TATGAACAGGAAGTCTGTGTGTTTACAACCAAGGAATGCATGCAGTGGTTCAGTCTGA 52403
QY 875 ulleValIleLeuLeuGlnCysLeuProAspGluLysCysLeuProLysAspIlePheAs 895
Db 52404 GATAGTCATTCTTACAGTGTCTACCGATGGAAGTGTTCGCAAGAGATATCTTTAG 52463
QY 895 nHisPheValGlnLeuTyrArgAspAlaLeuAlaGlyAsnValValSerAsnLeuGlyHi 915
Db 52464 TGACTTTGTGCAACTTTATCGGATCT-CTGGCAGGAATGTCGTGGCACTTTGGGACA 52522
QY 915 sSerPhePheSerGlnSerPheLeuGlySerLysGluHisGlyGlyPheLeuTyrValTh 935
Db 52523 TTCTCTTCTCAGTCAAGTCTCTCAGCAGTAAAAAACATCGTGGATTTTATATGTGAC 52582
QY 935 rSerThrTyrGlnSerLeuGlnAspLeuValLeuProThrProTyrProTyrLeuPheGlyI 955
Db 52583 ATCTACCTACAGTCACTGCAAGACCTAGTACTCCCAACCCACCTTACTTGTGGGAT 52642
QY 955 eleuileGlnLysTyrGluThrProTyrAlaLysValPheProIleArgLeuMetLeuAr 975
Db 52643 TCTTATCCAGAAATGGGAAACTCTTGGCTTAAAGTGTTCCTATCCATCTGATGTTGAG 52702
QY 975 gluGlyAlaGluTyrArgLeuTyrProCysProLeuPheSerValArgPheArgLysPPr 995
Db 52703 ACTTGGAGTGAATATCCACTTTATCCATGGCCACTAGTATGATTCAGATTTCGGAAGCC 52762
QY 995 oLeuPheGlyGluThrGlyHisThrIleMetAsnLeuLeuAlaAspPheArgAsnTyrGl 1015
Db 52763 ATTATTGGAGAGACGGGGCATACCATCATGAATCTTCTTCAGACCTTCAGAAATACCA 52822
QY 1015 nTyrThrLeuProValValGlnGlyLeuValValAspMetGluValArgLysThrSerIl 1035
Db 52823 GTATACCTTGGCAGTAGTTCAAGGTGTGGTGGTGTATGGAAGTTTGGAAACACCAT 52882
QY 1035 elysileProSerAsnArgTyrAsnGluMetMetLysAlaMetAsnLysSerAsnGluHi 1055

```

Db 52883 AAAAAATTCACGACACAGATAAAATGAGATGATGAAAGCCACGAAACAAAGTCCATGACGA 52942

Qy 1055 sValleuValaglyGlyAlaCysPheAsnGluLysAlaAspSerHisLeuValCysValG1 1075

Db 52943 TATCTTGGCAGAGGTGCTCTCTCAATGAAAGGACAGACTCTCATCTTGTGTGTA--- 52999

Qy 1075 nAsnAspAspGlyAsnTyrglnThrGlnAlaIleSerIleHisAsnGlnProArgLysVa 1095

Db 53000 -----AGCTATCAGACCCAGGCTATCAGTATTACAAATCAGCCAGAAAAGT 53046

Qy 1095 lThrGlyAlaSerPhePheValPheSerGlyAlaLeuLysSerSerSerGlyTyrlauAl 1115

Db 53047 GACTGGTGCCAGTCTCTTTGTTCAGTGGCACTCTGAAATACCTCTTCTGGAACCTTGC 53106

Qy 1115 aLysSerSerIleValGluAspGlyValMetValGlnIleThrAlaGluAsnMetAspSe 1135

Db 53107 CAAGTCCAGTATTGTGAAGATGGGTATGTGTCAGATCACTGCAGAAACATGGATTG 53166

Qy 1135 rLeuArgGlnAlaLeuArgGluMetLysAspPheThrIleThrCysGlyLysAlaAspAl 1155

Db 53167 CTTGAGGAGGCACTGTGAGAGATGAAGGACTTCACCATCACCCGTCGGGAGGCGCACGC 53226

Qy 1155 aGluGlnProGlnGluHisIleHisIleGlnTrpValAspAspLysAsnValSerLy 1175

Db 53227 AGAGATCCCGAGGAGACATCCACCCATGGTGTGATGATGACAGAACCTTAGCAA 53286

Qy 1175 sGlyValValSerProIleAspGlyLysSerMetGluThrIleThrAsnValLysIlePh 1195

Db 53287 GGGGCTGTAAAGTCTATAGATGGGAAGTCCATGGAGACTATAACAAATGTGAAGATATT 53346

Qy 1195 eHisGlySerGluTyrlsAlaAsnGlyLysValIleArgTrpThrGluValPhePheLe 1215

Db 53347 CTATGGATCAGAATGTAAGCAATGGAAGATGAAAGTCAATGAGTGGACAGAGGTGTTTCC 53406

Qy 1215 uGluAsnAspAspGlnHisAsnCysLeuSerAspProAlaAspHisSerArgLeuThrG1 1235

Db 53407 AGAAATCTCCACGAGCACATTCCTCAGTGATCCTGCAGATCACAGTAGATTGACTGA 53466

Qy 1235 uHisValAlaLysAlaPheCysLeuAlaLeuCysThrGlnLeuLysLeuLysGlyAs 1255

Db 53467 GCATGTTTACCAAGGCTTTTGTCTGTCTCTGCTCCTCACCTGAAGCTTCTGAAGGA 53526

Qy 1255 pGlyMetThrLysLeuGlyLeuArgValThrLeuAspSerAspGlnValGlyTyrglnAl 1275

Db 53527 TGGATGACCAAACTGGGACTACGTGTGACACTTGTATTCAGATCAGTCTGGCTATCAAGC 53586

Qy 1275 aGlySerAsnGlyGlnHisLeuProSerGlnTyMetAsnAspPheAspSerAspLeuVa 1295

Db 53587 AGGGAATAATGGCCAGCCCTTCCCTCGCACTGCATGAATGATCTGCACAGCGGCTTGGT 53646

Qy 1295 lLysMetIleHisGlyGlyAlaCysGlnLeuSerGluGlyProValValMetGluLeu11 1315

Db 53647 GCGGTGATTCATGGAGGGGCTGTCTGCTCAGTCTGAGGGCTCTGTCTGATGGAACATCAT 53706

Qy 1315 e 1315

Db 53707 T 53707

RESULT 10

AL671909/c

LOCUS Mouse DNA sequence from clone RP23-446F20 on chromosome 4, complete sequence.

DEFINITION

ACCESSION AL671909

VERSION AL671909.8 GI:22797964

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. Brown, J. 1 (bases 1 to 55802)

AUTHORS Brown, J.

TITLE Direct Submission

JOURNAL

COMMENT

Submitted (06-SEP-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Sep 12, 2002 this sequence version replaced gi:21212224. clones. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-446F20 is from the RPI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACes.6

Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

FEATURES

source

1. 85802
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-446F20"
/clone_lib="RPC1-23"

ORIGIN

Alignment Scores:

Pred. No.: 1,06e-152 Length: 85802
Score: 2471.50 Matches: 482
Percent Similarity: 84.37% Conservative: 47
Best Local Similarity: 76.87% Mismatches: 71
Query Match: 35.36% Indels: 27
DB: 10 Gaps: 6

US-09-744-167-2 (1-1323) x AL671909 (1-85802)

Qy 1 MetTrpIleAspGluAsnAlaValAlaGluAspGlnLeuIleLysArgAsnTyrsrTp 20
|||||
Db 60423 ATGTGGATTGATGAAATCTGTGACAGAGGACCAACTAATTAAGAGAACTATATACG 60364
|||||
Qy 21 AspAspGlnCysSerAlaValGluValGlyGluLysCysGlyAsnLeuAlaCysLeu 40
|||||
Db 60363 GATGATCAATTAGTGCCCTTGAAGTGGAGAGAGAGAACTGGAAGCTCACTTGTCTG 60304
|||||
Qy 41 ProAspGluLysAsnValValValValAlaValMetHisAsnCysAspLysArgThrLeu 60
|||||
Db 60303 CCAGACGAGAAGAAATCTTCTGTTGTAGCTGTGATGATGATGATGATGATGATGATG 60244
|||||
Qy 61 GlnAspAspLeuGluAspCysAsnAsnTyrsrAsnSerGlnSerLeuMetAspAlaPheSer 80
|||||
Db 60243 CAAAGCGATTTCAGGATTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 60184
|||||
Qy 81 CysSerLeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGluSer 100
|||||
Db 60183 TGTTCACTGGATAATCAGACTAGACAACTGATCAATTAATTAATTAATTAATTAATTAAT 60124
|||||
Qy 101 ThrGluLysAspMetAsnSerGluLysGlnMetAspProLeuAsnArgProLysThrGlu 120
|||||

60123	Db	ACGGAAAAAGGTATAAACTCGGAGAAACAAATGGATGCTTTTGAATAAACCAAAACACGAG	60064
121	Qy	GlyArgSerValAasnHisLeuCySProThrSerSerAspSerLeuAlaSerValCysSer	140
60063	Db	AGGATTCTGTGAATCATCTGTGGGGCTTCATCTAACAGTGCACACGATATCTCTTCC	60004
141	Qy	ProSerGlnLeuLysAspAspGlySerIleGlyArgAspProSerMetSerAlaIleThr	160
60003	Db	CCITTCAAAATTGAAGACGGTGAGAAATGTGGGTAGAGACCCATCTACATCTTACAGTTACA	59944
161	Qy	SerLeuThrValAspSerValIleSerSerGlnGlyThrAspGlyCysProAlaValLys	180
59943	Db	AGCTTAGAGTTAAAT-----TCATCCAGGGATGGATGGAGGGCTGCCATTAA	59893
181	Qy	LysGlnGluAsnTyrrileProAspGluAspLeuThrGlyLysIleSerSerProArgThr	200
59892	Db	CAGCAAGGAAATATTATGCCAGATGAGGATCTTCTGGCATGAACAGAGCTTCTCCAGGACA	59833
201	Qy	AspLeuGlySerProAsnSerPheSerHisMetSerGluGlyIleLeuMetLysLysGlu	220
59832	Db	GACCTAGGAGATCTCCAACTCCITTTCCCACTCAAGTGGAGAGCTTTATAAAAAACAGAG	59773
221	Qy	ProAlaGluGluSerThrThrGluGluSerLeuArgSerGlyLeuProLeuLeuLeuLys	240
59772	Db	CCAGCTGAGGAGGACAGCTGAAGACTCCCTCCCATCTGATTTATCTTGAATCTCAAG	59713
241	Qy	ProAspMetProAsnGlySerGlyArgAsnAsnAspCysGluArgCysSerAspCysLeu	260
59712	Db	CCGGATACTCCTCTTTATCTTGAAGGGACAAT---TGGCAACAGTCTTCAGACTGCCTT	59656
261	Qy	ValProAsnGluValArgAlaAspGluAsnGluGlyTyrGluHisGluThrLeuGly	280
59655	Db	GGGTCTAGTGAAACTAGACTGATGAAGTAGAGGT-----	59620
281	Qy	ThrThrGluPheLeuAsnMetThrGluHisPheSerSerGluSerGlnAspMetThrAsnTrp	300
59619	Db	-----AATGACTCTCAG---ATGACTAATTTGG	59596
301	Qy	LysLeuThrLysLeuAsnGluMetAsnAspSerGlnValAsnGluGluLysGluLysPhe	320
59595	Db	AAATTAACGAAGCTAAATCAGATGAGTGATAGCCAAAGTAATGAAGAAAAATCAATGGCT	59536
321	Qy	LeuGlnIleSerGlnProGluAspThrAsnGlyAspSerGlyGlyGlnCysValGlyLeu	340
59535	Db	TTACAGAGTAATCAACTCGAGACACTAAT---AGTGGAGGCGAGTGTGTGGAATG	59482
341	Qy	AlaAspAlaGlyLeuAspLeuLysGlyThrCysIleSerGluSerGluGluCysAspPhe	360
59481	Db	GCAGATCTTGATCTAGATTCAAAGGAAGCTTCGATGAATGAGAGTCAAGGATATGATTTC	59422
361	Qy	SerThrValIleAspThrProAlaAlaAsnTyrrLeuSerAsnGlyCysAspSerTyrrGly	380
59421	Db	TCCACTGTTAATGATGCACCAGCAGCAAAATCTCTATCTAATAGTTGTGATTTCCTATGGA	59362
381	Qy	MetGlnAspProGlyValSerPheValProLysThrLeuProSerLysGluAspSerVal	400
59361	Db	ATGCAGAGCCCAATTGTATCTTTTGTTCAAAAGACATTACCTCTCAAAGAGAGATTACGTA	59302
401	Qy	ThrGluGluLysGluIleGluGluSerLysSerGluCysTyrrSerAsnIleTyrrGluGln	420
59301	Db	ACAGAGAAAGAAAGAAATAGAGNAGCAGCAAGTCAGATGCTATTCAAATATTATTAAGAACAG	59242
421	Qy	ArgGlyAsnGluAlaThrGluGlySerGlyLeuLeuLeuAsnSerThrGlyAspLeuMet	440
59241	Db	AGAGGAAACAGAGAAATGCAGAGGGAGTGGACTGCTTTTAAACAGCACTGGGAAATGTAATG	59182
441	Qy	LysLysAsnTyrrLeuHisAsnPheCysSerGlnValProSerValIleGluGlnSerSer	460
59181	Db	AAGAAAAAATTTTATCATTAATTTCTGTAGTCAAAATCCCATCAGTACTTTGGGCATCTTCC	59122
461	Qy	ProLysValValAlaSerLeuProSerIleSerValProPheGlyGlyValaArgProLys	480

59121	CCCAAG---ATAGCAAACCTGC	AATCTATTAGTGTTCCTTTGGTGGTGC	CAAGACCAAG	59065
Qy	481	GlnProSerAsnLeuLysLeuGlnIleProLysProLeuSerAspHisLeuGlnAenAsp	500	
Db	59064	CAACCTCTAATCTTAAACCTTCAGATTCCAAAGCCATTATCAGACCATTTACAAATGAC	59005	
Qy	501	PheProAlaAsnSerGlyAsnAsnThrLysAsnLysAsnAspIleLeuGlyLysAlaLys	520	
Db	59004	CTTCCTGCAACCAATGGAATATATGATAAAACAAAAATGATGTCTTGGGAAGACCAAG	58945	
Qy	521	LeuGlyVLAsnSerAlaThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal	540	
Db	58944	CTAGGGGAGAACTCAGCAGTCAACGCATCGAGTCGCAGCTCGGAAACATCTCTGTAACT	58885	
Qy	541	AspThrAsnGlyGluHisIleuGlySerTyrgluAlaGluIleSerThrArgProCysLeu	560	
Db	58884	GATACAAATGGGAAACATTTAGAAACCTTAGAGCTTAGGGCTGAGATCTCCAGATGACCATGCCTT	58925	
Qy	561	AlaLeuAlaProAspSerProAspAsnAspLeuArgAlaGlyPheGlyIleSerAla	580	
Db	58824	GCCTAGCTCCAGATAGTCCAGATATGATCTCCGAGCTAGTCAGTTGGATTCTGCA	58765	
Qy	581	ArgLysProPheThrThrLeuGlyGluValAlaProValThrValProAspSerGlnAla	600	
Db	58764	AGAAAGCCATTTACCAACCTTAGTGAAGTGGCACCAGTAGTGGTACCAAGATTTCCAGGCT	58705	
Qy	601	ProAsnCysMetLysCysGluAlaArgPheThrPheThrLysArgArgHisCysValArg	620	
Db	58704	CCAACTGCATGAAGTGTGAAGCTCGTGTATCATTCACCAAAAGAGAGCATCCTGCAGA	58645	
Qy	621	AlaCysGlyLysValPheCys	627	
Db	58644	CGGTGTGGGAAGCTAAGTTGC	58624	
RESULT 11				
AC095988/c				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				

Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Mines, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaekemele, O., Okwunigbo, G., Olarunpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., PrannKoch, C.,
Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puzao, M., Qureshi, J., Rachlin, E., Reeves, K., Regier, M.A., Reight, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,
Sneedi, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vega, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H. O.,
Weinstock, G. and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:24818127. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GECU
Center clone name: CH230-29P9
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 248170 bases at least Q40
Consensus quality: 250604 bases at least Q30
Consensus quality: 252206 bases at least Q20
Estimated insert size: 262793; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```

Db	181881	TCCTTCAAAATTAAGGACAAATGAAATGTTAGTAGAGACCCATCTACATCTACAGCTACA	181822	Db	180882	CTTCTCGCAACAATGGAATAATAGTAAACAAATGATATCTCTGGAAAGCAAAA	180823
Qy	161	SerLeuThrValAspSerValIleSerSerGlnGlyThrAspGlyCysProAlaValLys	180	Qy	521	LeuGlyGluAsnSerAlaThrAsnValCysSerProSerLeuGlyAsnIleSerAsnVal	540
Db	181821	AGCTTAGCAATTAAT-----ACATCCAGGGAATGATGAGGCGCTGCAATTA	181771	Db	180822	CTAGGGGAAAACCTCAGCAGTCAGTGCAGTCACTTTGGGAAAACATCTCTGTGACT	180763
Qy	181	LysGlnGluAsnTyrIleProAspGluAspLeuThrGlyLysIleSerSerProArgThr	200	Qy	541	AspThrAsnGlyGluHisLeuGluSerTyrGluAlaGluIleSerThrArgProCysLeu	560
Db	181770	AAACAAGAGGATTATATGCTGATGAGGACCTTTCTGGCGCAACACAGCTCTCTAGGACA	181711	Db	180762	GATACAAATGGGAAACATTTAGAAGTTTCGAGGCTGCATCTCCAGTAGCAATGCTT	180703
Qy	201	AspLeuGlySerProAsnSerPheSerHisMetSerGluGlyIleLeuMetLysLysGlu	220	Qy	561	AlaLeuAlaProAspSerProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAla	580
Db	181710	GATCTAGGAGCTCAAACTCTCTTCCACTCAAGGAGGAGCTCTTAACAAAACACAG	181651	Db	180702	GCATTATCTCCAGATAGTCCAGATAATGATCTCAGAGCTGCTCAATTGGAAATCTTGA	180643
Qy	221	ProAlaGluGluSerThrThrGluGluSerLeuArgSerGlyLeuProLeuLeuLys	240	Qy	581	ArgLysProPheThrThrLeuGlyGluValAlaProValTrpValProAspSerGlnAla	600
Db	181650	CCAGCAGAGGAGAGGACAGCTGAAGACCCCTCCATCTGGTTATCTTTGAATCTCAAG	181591	Db	180642	AGAAAGCCATTTTACACCTCGGTGAGTGCCAGCATGGTACCAAGATCCAGAGCT	180583
Qy	241	ProAspMetProAsnGlySerGlyArgAsnAsnAspCysGluArgCysSerAspCysLeu	260	Qy	601	ProAsnCysMetLysCysGluAlaArgPheThrPheThrLysArgArgHisCysArg	620
Db	181590	CCAGATACCTCTGCTTTATGTGGAGGACAAAT---TGCGAACCATCTCAGACTGCTT	181534	Db	180582	CCGAATCTGCATGCACTGAGTGTGAGCTCGTTTACATTCACCAAGAGGAGCATCATTCGAGA	180523
Qy	261	ValProAsnGluValArgAlaAspGluAsnGluGlyTyrGluHisGluGluThrLeuGly	280	Qy	621	AlaCysGlyLysValPheCys 627	
Db	181533	GTGCTAGTGAGACTGAGACTGATGGAGATGGAGT-----	181498	Db	180522	GCATGTGGAGAGTAAATGTC 180502	
Qy	281	ThrThrGluPheLeuAsnMetThrGluHisPheSerGluSerGlnAspMetThrAsnTrp	300	RESULT 12			
Db	181497	-----AAGGACTCTCAGACGACTCACTGG	181474	AC118106	201330 bp	DNA	linear
Qy	301	LysLeuThrLysLeuAsnGluMetAsnAspSerGlnValAsnGluGluLysGluLysPhe	320	LOCUS	HTG 13-NOV-2002		
Db	181473	GACTCAGCAAGCTAAATGAGTGGATGATCCCAAGCAGAGTGAAGAAATCAAGTGGT	181414	DEFINITION	***, 2 unordered pieces.		
Qy	321	LeuGlnIleSerGlnProGluAspThrAsnGlyAspSerGlyGlnCysValGlyLeu	340	ACCESSION	AC118106.5	GI:24941760	
Db	181413	TTACAGATCAGCCAACTGAGACCT-----CAGAGTGCAGGAGAGTGTGTAGAATG	181360	VERSION	HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.		
Qy	341	AlaAspAlaGlyLeuAspLeuLysGlyThrCysIleSerGluSerGluGluCysAspPhe	360	KEYWORDS	Rattus norvegicus (Norway rat)		
Db	181359	CGCAGATTCCAGATCTAGATTTCAAAGGAACCTTGGTGAATGAAGTGAAGATATGATTC	181300	SOURCE	Rattus norvegicus		
Qy	361	SerThrValIleAspThrProAlaAlaAsnTyrLeuSerAsnGlyCysAspSerTyrGly	380	ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Db	181299	TCCACTGTAGTGATGCACAGCAGCAAAATCTCTGTCTAATAGTCTGATCTCTATGGA	181240	REFERENCE	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;		
Qy	381	MetGlnAspProGlyValSerPheValProLysThrLeuProSerLysGluAspSerVal	400	1	Rattus.		
Db	181239	ATGCAGAGCCCAATGTATCTTTGTTCCAAAGACTTTACCTCCAAAGAGATTCAGTA	181180	1	(bases 1 to 201330)		
Qy	401	ThrGluGluLysGluIleGluSerLysSerGluCysTyrSerAsnIleTyrGluGln	420	Muzay, D. Marie, Metsker, M. Lee, Abramzon, S., Adams, C., Alder, J.,			
Db	181179	ACAGAAGAAAAGAAATAGAAGAGCAAAATCAGATGCTACTCAATCTTTATGAACAG	181120	Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,			
Qy	421	ArgGlyAsnGluAlaThrGluGlySerGlyLeuLeuLeuAsnSerThrGlyAspLeuMet	440	Anylebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,			
Db	181119	AGAGGAAATGAGAACACAGAGGGAATGAGCTGCTTTTAAACAGCACTTGAACATAATG	181060	Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,			
Qy	441	LysLysAsnTyrLeuHisAsnPheCysSerGlnValProSerValLeuGlyGlnSerSer	460	Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,			
Db	181059	AAGAAAAATTTATTTACATAATTTCTGTAGTCAATCCCATCTGCTACGGCAATCTCC	181000	Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,			
Qy	461	ProLysValAlaSerLeuProSerIleSerValProPheGlyGlyAlaArgProLys	480	Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,			
Db	180999	CCCAAG---ATAGCAAACTGCAGCTATTAGTGTCTCTTTTGGTGGCAAGACCCAAAG	180943	Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,			
Qy	481	GlnProSerAsnLeuLysLeuGlnIleProLysProLeuSerAspHisLeuGlnAsnAsp	500	Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,			
Db	180942	CAACCTCTTAATCTTAACTTCAAGTCCCAAGGCAATATCAGACCAATTTACAAAATGAC	180883	Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,			
Qy	501	PheProAlaAsnSerGlyAsnAsnThrLysAsnLysAsnAspIleLeuGlyLysAlaLys	520	Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,			

Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rokey, I., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waidron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

TITLE JOURNAL

REFERENCE
2 (bases 1 to 201330)
Worley, K.C.

TITLE JOURNAL

Submitted (13-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 201330)
Rat Genome Sequencing Consortium.

REFERENCE AUTHORS TITLE JOURNAL

Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 13, 2002 this sequence version replaced gi:23194893.

COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly ('a contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GVBD

Center clone name: CH230-324B23

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 188149 bases at least Q40

Consensus quality: 190201 bases at least Q30

Consensus quality: 191458 bases at least Q20

Estimated insert size: 194406; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 198699: contig of 198699 bp in length

* 198700 198799: gap of unknown length

* 198800 201330: contig of 2531 bp in length.

* Location/Qualifiers

FEATURES

```

source          1..201330
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-324B23"
misc_feature    1..1448
/note="wgs end extension
clone_end:Sp6"
misc_feature    1920..4036
/note="wgs end extension
clone_end:Sp6"
misc_feature    complement(6631..7552)
/note="clone boundary
clone_end:Sp6
site_end:Sp6
end_sequence=BZ238666"

ORIGIN
Alignment Scores:
Pred. No.:      1,69e-124      Length:      201330
Score:          2056.00        Matches:     416
Percent Similarity: 74.80%    Conservative: 53
Best Local Similarity: 66.35%  Mismatches:  131
Query Match:     29.41%       Indels:      28
DB:              2            Gaps:        5

US-09-744-167-2 (1-1323) x AC118106 (1-201330)
QY      1 MetTPIIAspGluAsnAlaValAlaGluAspGlnLeuIlelyAsnTyrSerTrp 20
Db      34915 ATGTGGATTGATGAATAATCTGTGCAGAGAACCAACTAATTAAGAGAACTACATCAG 34856
QY      21 AspAspGlnCysSerAlaValGluValGlyGluLysLysCysGlyAsnLeuAlaCysLeu 40
Db      34855 GATGATCAATTTAGTGGCGTTTGAAGTGGAGAGAGAAACGTGGGAGCCTCACTGTCTG 34796
QY      41 ProAspGluLysAsnValLeuValValAlaValMetHisAsnCysAspLysArgThrLeu 60
Db      34795 CCAGACGAGAGAAGATTTCTGTTGTGTAGTGTGCATGCATAACTGTGATAAAGGACACTG 34736
QY      61 GlnAsnAspLeuGlnAspCysAsnAsnTyrAsnSerGlnSerLeuMetAspAlaPheSer 80
Db      34735 CAAAGTGATTGACAGGATTTGTAATTAATATATATATATCAATCCCTCATGGATTTCTTTAGC 34676
QY      81 CysSerLeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGluSer 100
Db      34675 TGTTCACCTGGATAATGAACACTAGACAACTGATCAATTTAGTTTATGATGAATGGGTCC 34616
QY      101 ThrGluLysAspMetAsnSerGluLysGlnMetAspProLeuAsnArgProLysThrGlu 120
Db      34615 GCAGAAAAAGGTATTAACCTCAGAGAAACAAATGGATGCTTTGAAATAAACCAAAATCAGAG 34556
QY      121 GlyArgSerValAsnHisLeuCysProThrSerSerAspSerLeuAlaSerValCysSer 140
Db      34555 AGGGATTCTCTTAATCACTCTGCGCTTCAGCTTAACAGTCAACAGTCACTCTCTTCC 34496
QY      141 ProSerGlnLeuLysAspAspGlySerIleGlyArgAspProSerMetSerAlaIleThr 160
Db      34495 TCTTCACAAATTAAGACAACTGAAATGTAGTAGAGACCCATCATCTACAGCTACA 34436
QY      161 SerLeuThrValAspSerValIleSerSerGlnGlyThrAspGlyCysProAlaValLys 180
Db      34435 AGCTTAGCAATTAAT-----ACATCCCAAGGAATGGATGGAGGCGCTGCCATTAAA 34385
QY      181 LysGlnGluAsnTyrIleProAspGluAspLeuThrGlyLysIleSerSerProArgThr 200
Db      34384 AAACAAGAGGATTATATGCTTCATCAGGACCTTTCTGGCGCAACACAGCTCTTCTTAGGACA 34325
QY      201 AspLeuGlySerProAsnSerPheSerHisMetSerGluGlyIleLeuMetLysLysGlu 220
Db      34324 GATCTAGGAGGCTCAAACTCTTTTCCCACTCAAGGAGGAGGCTCTTAACAAAAACAGAG 34265
QY      221 ProAlaGluGluSerThrThrGluGluSerLeuArgSerGlyLeuProLeuLeuLys 240

```

[illegible]

Db 33257 AGANNNNNNNNNNNNNNNTGGTGAGGTGGCCACGATGATGGTACCAGACTCCAGCCT 33139

Qy 601 ProAsnCysMetLysCysGluAlaArgPheThrLysArgArgHisHisCysArg 620
CGAACTGCATGAAGTGTGAAGTCGTTTACATTACCAAAGAGGAGCATCTTGCAGA 33138

Db 33197 CCGAACTGCATGAAGTGTGAAGTCGTTTACATTACCAAAGAGGAGCATCTTGCAGA 33138

Qy 621 AlaCysGlyLysValPheCys 627
|||||
|||

Db 33137 GCATGTGGGAAGGTAAATGCG 33117

RESULT 13

LOCUS BC042669 5480 bp mRNA linear ROD 07-OCT-2003

DEFINITION Mus musculus RIKEN cDNA B130024H06 gene, mRNA (cDNA clone MGC:28649 IMAGE:4235024), complete cds.

ACCESSION BC042669

VERSION BC042669.1 GI:27503685

KEYWORDS MGC.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 5480)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klautner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altshuler,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,A.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.J., Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carinci,P., Prange,C., Raja,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Hellon,E., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Kravinsky,M.I., Skalska,U., Smalley,D.E., Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002) 22388257

TITLE Direct Submission
Submitted (02-JAN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
Contact: MGC help desk
Email: gcaps-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: ang@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowitz, C.R., Speed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: http://image.llnl.gov Series: IRAK Plate: 37 Row: k Column: 18
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA Gi: 27734995.

FEATURES

Location/Qualifiers
 1..5480
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="MGC:28649 IMAGE:4235024"
 /tissue type="kidney, normal, 5 month old male mouse."
 /clone_lib="NCI CGAP_Kid14"
 /lab_host="DH103"
 /note="vector: pCMV-SPORT6"
 1..5480
 /gene="B130024H06Rik"
 /note="synonym: B130031I15"
 /db_xref="locusID:218441"
 /db_xref="MGI:2145181"
 150..4736
 /codon_start=1
 /product="RIKEN cDNA B130024H06"
 /protein_id="AAH42669.1"
 /db_xref="GI:27503686"
 /db_xref="LocusID:218441"
 /translation="MDSYFKAAYSGLDKLLDDPEONPKODYLPDAYAFNCQSVSES
 ASPLALLSDQRICSTCASSECCEDANETFLGKIHEGLTRSRNEKXNVAGDULLSS
 VDASTDEIQSCMRRCSEKFCVDDISDMGNLVKATNEEDIKQLLPDDPKSSADTLIT
 LDSSVSEALTVSSVDCGNAVREEQNNINAGNKNDISIKELGVKVDMLDFDCKYN
 RTENLKDIISNELETVDYFDMPSVMEQSESMGNTKNDPQYKLPCELLKDDGLAE
 KVAVANVTECLERGGSNTIAMPKLPENEGISPSDPAKSKDENFKLPDPLOENRTS
 VFMKQYKEDSRNLDKNDNDIVHVGDDVPVPSLCSLSGSLSGSLGSLHNNHSDILP
 PISEGQNNADATHEIKQSDVLDGETDLSKKETCHSIPLQPVNEKCKGKVEVREM
 VISGESSEDEDSAAAGSPVALSAASVPEAPGPCGELTFPSSMDGQELDYFNID
 ESMKSGLISDAEIDAFLEKQCSLNTSMAGENVDLSQMQNQLTMKGLHENDAGDI
 YFNAAGAAGENGPCVNCETSKDENENGLSISGEGKAIPTERELSAQVPIDEILPV
 PSITQVAGGARPKQLSLPPGTRSKENLKPVDVDPSEPCATANATVSTCSADHI
 PDQVSNVNTIDTESFEDGSGFVTANKDSLLENKRESLVLGOKPTWVPDSEAPN
 CMNCQVATKRRHRCACQKVCVGNCKKLOYLEKEARVCVICYETINKAQAF
 ERMSPGSGCLKNSNECATDPLQETQTSSTPSPTLPIALSALKNPVEGPCSKQK
 RWFADGILNGEVADTITKLSSGSKRCDSPVLDPDVPFMINKVORTHSPEETKPN
 GLGDIIEISQVCTHAFVERLPGNTGTEGLPMGPFTLEDDVFDVDSPEPTTW
 PANGPLVASISDRLVCSVAKVCNNISLLPDDDIGLPLLATSGEDSGVPVYQVRP
 SHEQIILLLEGEPPTAFVNLNANLVNKLVSSEKYSWFSTNGHLGQAEITVL
 LQCPNEDVPKDFILFIYKDALGKVIENLDNLTFTFESLNSKHGSGPLFIPT
 FONDGHPVPSRPLCGILIQLEIPKAPVPMELMLRLGAEYKAPALTSVGRKBP
 LFGEIGHTNMLLDNRNYQVTLNIDQLLIHMEGSKCIKPRKYSVDMVKVHSN
 EHVISGASFTEADSHLVQSDGVYQVQANSATQPRKVTGASFFVFNAGLTKSSG
 FLAKSIVELGQVQIPTETMEGLRLAREQKDFRQCQKVDADRLREYVDCWDSSE
 ERKNKGVSSVDGMSVGFPESEKIKLETDETEKTVKTEVFYFKDDQDISLSSY
 QFAKEIAVACSAALCPHLRLKSNRNKIGLRYSIDTDMVEFQAGCGQLLPQHLYND
 LDSALIPVINGTSSNSLPLEILELAPFLENLSE"
 2337..2531
 /note="FYVE; Region: FYVE zinc finger. The FYVE zinc
 finger is named after four proteins that it has been found
 in: Fab1, YOTB/ZK632.12, Vac1, and EEA1. The FYVE finger
 has been shown to bind two Zn++ ions. The FYVE finger has
 eight potential zinc coordinating cysteine positions. Many
 members of this family also include two histidines in a
 motif R-HHC+XCG, where + represents a charged residue and
 X any residue. We have included members which do not
 conserve these histidine residues but are clearly related"
 /db_xref="CDD:pfam01363"

misc_feature

ORIGIN

Alignment Scores:

Pred. No.: 3,86e-126 Length: 5480
 Score: 2047.00 Matches: 523
 Percent Similarity: 49.56% Conservative: 208
 Best Local Similarity: 35.46% Mismatches: 481
 Query Match: 29.28% Indels: 264
 DB: 10 Gaps: 40

US-09-744-167-2 (1-1323) x BC042669 (1-5480)

Qy 3 IleAspGluAsnAlaValAlaGluAspGlnLeuIleLysArgAsnTyrSerTrpAspAsp 22

Db 630 TTGGACTCGTCTTCAGTTTCAG----- 653
 Qy 23 GlnCysSerAlaValGluValGlyGluLysLysCysGlyAsnLeuAlaCysLeuProAsp 42
 Db 654 -----GCTCTCACTCTCTCTCAGTAGACTGTGGTAGTAGTCTGTCTCAGAGAAGAG 704
 Qy 43 GluLysAsnValLeuValAlaValMetHisAsnCysAspLysArgThrLeuGlnAsn 62
 Db 705 CAGATATACATC-----AACGCTGGTATATAAACACAGAGATATT 743
 Qy 63 AspLeuGlnAspCysAsnAsnTyrAsnSerGlnSerLeuMetAspAlaPheSerCysSer 82
 Db 744 AGTATCAAGAATTTGGTGTAAAGTAGACATGGCACTTTTTCAT-----TCTTGTATAA 797
 Qy 83 LeuAspAsnGluAsnArgGlnThrAspGlnPheSerPheSerIleAsnGluSerThrGlu 102
 Db 798 TATAATAGAACAGAAATTTAAAGAGATAAAATATTCTTAATGAGTTAGTAGACAGTTGAT 857
 Qy 103 LysAspMetAsnSer-----GluLysGlnMetAspProLeuAsnArgProLysThrGlu 120
 Db 858 TTGACATCCCATCTGTTTGTATGGAACAAAGTTCTGAAATGCTCTAATACCAAGATAAC 917
 Qy 121 GlyArgSerValAsnHisLeuCysProThrSerSerAspSerLeuAlaSerValCysSer 140
 Db 918 CCACAATACAGAGCGTCCATGT----- 941
 Qy 141 ProSerGlnLeuAspAspGlySerIleGlyArgAspProSerMetSerAlaIleThr 160
 Db 942 ---GAATTGTTAAAGATGATGCTGTTGGCAGAGAGAAAGTAGCTGCGCAGTCAAT 998
 Qy 161 -----SerLeuThrValAsp-----SerVal 167
 Db 999 AATACAGACTGTTTAAAG 1058
 Qy 168 IleSerSerGlnGlyThrAspGlyCysProAlaValLysGlnGluAsnTyr---Ile 186
 Db 1059 CAGAAATGAAGAGGATATCCCAAGTAGCCAGCTTCAAAAGATGAAAATTTCAAGTTA 1118
 Qy 187 ProAspGluAspLeu-----ThrGlyLysIle 195
 Db 1119 CQTGACTTTCCCTCAGGAAAAATAGAACTTCTGTGTTTCATGAAACAGACTGTAAAGAA 1178
 Qy 196 SerSerProArgThrAspLeuGlySerProAsnSerPheSerHisMetSerGluGlyIle 215
 Db 1179 GACTCAAGAAATTTAGACCTTAAAGATAATAATGATATAGTCCATGTTTCA----- 1229
 Qy 216 LeuMetLysLysGluProAlaGluSerThrThrGluLysSerLeuArgSerGlyLeu 235
 Db 1230 -----GGTGATGATGTGCCCTTCACTGCTCTCTTTCATGCTGGGCTTTG 1280
 Qy 236 ProLeuLeuLysProAspMetProAsnGlySerGlyArgAsnAsnAspCysGluArg 255
 Db 1281 -----TGTGGATCGTTAAATTCACAATAAT-----GAA 1307
 Qy 256 CysSerAspCysLeuValProAsnGluValArgAlaAspGluAsnGluGlyTyrGluHis 275
 Db 1308 CACAGTATATTTTACCTCCTAATGAGTCTGAAGCCAGACAAATGATGAGTACTATA 1367
 Qy 276 GluGluThrLeuGlyThrThrGluPheLeuAsnMetThrGluHisPheSerGluSerGln 295
 Db 1368 CATGAAGAGATACAGAAGAGTGTATGTTTAGATGGGGAGACAGATCTTTCAAAAGAGAA 1427
 Qy 296 AspMetThrAsnTrpLysLeuThrLysLeuAsnGluMetAsnAspSerGlnValAsnGlu 315
 Db 1428 AGCTGTAGAGCATATTTCTTCAGCCAGTAAATGAAAGAAAGGGAGGAGGAAAGGTTGAG 1487
 Qy 316 GluLysGluLysPheLeu-----GlnIleSerGlnProGluAspThrAsnGly--- 331
 Db 1488 GTTGAGGATGATGTAATCAGTGTGAGTCTTTGGAGTCCCTCTGAGGACCCAGCTCTGCT 1547
 Qy 332 -----AspSerGlyGln 336

1548 CGACGAGCTGGTCTCGGTGGCACTTTCTGCTGCAAGTGTCCAGAGGCTCTGGTCCA 1607
337 CysValGlyLeuAlaAspAlaGlyLeuAspLeuLysGlyThrCysIleSerGluSerGlu 356
1608 TGTGAAGCTCTCACTTTTCCAGCAGCTGATATGGATGGG-----CAA 1649
357 GluCysAspPheSerThrValIleAspThrProAlaAlaAsnTyrLeuSerAsnGlyCys 376
1650 CAGITAGATTACTTTAATATTGATGAAGCATGAGAGTGGCATACTAATTAGCGA-TGC 1708
377 AspSerTyrGlyMetGlnAspProGlyValSerPheVal-ProLysThrLeuProSerLys 396
1709 TGAACCTGATGCTTTCTGAAGGACACAGTGTCTTTCGAACCTCTAACCAATGTCTGCTGG 1768
396 sGlu-----AspSerValThrGluGluLysGluIle-----G1 407
1769 AGAAATGTAATGATTCCTCCAGTTGCAGATGAATCAGATCACTATGAAGAGCTACATGA 1828
407 uGluSerLysSerGluCysTyrSerAsnIle----- 417
1829 TGAAATCTCGGGATATATACTTCAATGCTGAAGCAGGAGCTCTGGAGAAATGGAGG 1888
418 -----TyrGluGlnArgGlyAsnGluAlaThrGluGlySerGlyLeuLeuAs 434
1889 TGTGGCAACTGTGAACAAGTGATAAAGAAACACACAAATAATATGTTTCCATAGG 1948
434 nSerThrGlyAspLeuMetLysLysAsnTyrLeuHisAsnPheCysSerGlnValProSe 454
1949 AGAAAGGCGCAATTCACATGAAGA----- 1976
454 rValLeuGlyGlnSerSerProLysValValAlaSerLeuProSerIleSerVal----- 472
1977 -GAGTTATCTGCTGCTGACGCTGATGAAGGATGAATTCAGTCCCGCATTAATAAC 2035
473 ----ProPheGlyAlaArgProLysGlnProSerAsnLeuLysLeuGlnIleProLys 491
2036 CCAAGCTGTTGGAGGGCTCGACCTTAAGCAG-----CTGCTGAGTCTTCCA-- 2081
491 sProLeuSerAspHisLeuGlnAsnAspPheProAlaAsnSerGlyAsnAsnThrLysAs 511
2082 -----CCAGGAACAGGAGTTCAAAGGAACGTAA 2110
511 nLysAsnAspIleLeu-----GlyLysAlaLysLeuGlyGluAsnSerAlaTh 527
2111 TAAGCCAGATGTTGGATGTGCCAGAAAGTGACCCATGACACAAATGCTACTGTGT 2170
527 rAsnValCysSerProSerLeuGlyAsnIleSerAsnValAspThrAsnGlyGluHisLe 547
2171 ATCCACTTGTCTGCAGATCATATACCTGATTTCTCAGGTTAGCTTCAACTCCAATTACAT 2230
547 uGluSerTyrGluAlaGluIleSerThrArgProCysLeuAlaLeuAla----ProAspSe 566
2231 TGAC---ATAGAAGTAATTTTGAAGATGGATCCAGTTTGTATACTGCAACACAGGATTC 2287
566 z---ProAspAsnAspLeuAlaGlyGlnPheGlyIleSerAlaArgLysProPheTh 585
2288 TCTGCTGAAAT-----AAGCGCAAGAAAGACTT 2317
585 rThrLeuGlyGluValAlaProValTrpValProAspSerGlnAlaProAsnCysMetLys 605
2318 GGTTTTGGCCAGAAACCACTACATGGTTCCTGATTCAGAAGCTCCAACTGTATGAA 2377
605 sCysgluAlaArgPheThrLysArgArgHisIleCysArgAlaCysGlyLysVa 625
2378 CTGCCAAGTCAAAATTCACCTTTACAAAACGGCGACACCCTGCCAGCATGTGGGAAGGT 2437
625 lPheCysAlaSerCysCysSerLeuLysCysLysLeuLeuTyrMetAspArgLysGluAl 645
2438 ATTTTGTGGTGTCTGTTAAACAGGAAATGTAAGCTTCAGTATCTAGAG--AAGGAAGC 2494
645 aArgValCysValIleCysHisSerValLeuMetAsnAlaGlnAlaTrpGluAsnMetMe 665
2495 AAGAGTATGTGTAATCTGCTATGAGACTATAATAAAGCTCAGGCATTTTGAAGAGTATG 2554

665 tSerAlaSerSerGlnSerProAsnProAsnAsnProAlaGluTyrCysSerThrIlePr 685
2555 GAGTCCAGGTGTTCTTGTCTTAAGTCTAATCATCTTAATGAA---TGTGCCACTGATCA 2611
685 oProLeuGlnGlnAlaGlnAlaSerGlyAlaLeuSerSerProProThrValMetVa 705
2612 GCCACTTCAGGAGACCCAAACATCC-----AGTACACCTTCACCTACAACTTTG-- 2660
705 lProValGlyValLeuLysHisProGlyAlaGluValAlaGlnProArgGluGlnArgAr 725
2661 -CCATCTTCAGCACTGAACCAACCAAAAGTTGAAGGACCATGTTCCAAAGAACAGAGAG 2719
725 gValTrpPheAlaAspGlyIleLeuProAsnGlyValAlaAlaAspAlaAlaLysLeu-- 744
2720 AGTATGGTTTGCAGATGGCATACTCCGAATGGTGAAGTTGCAGATACCTAGCAAAATTATC 2779
745 -----ThrMe 746
2780 ATCTGGAAGTAAAGATGTTCTGATGACTTTAGTCTCTCTTACTGATGTGCCACAAT 2839
746 t-----AsnGlyThrSe 750
2840 GATAAATAAGTGGATCGTACCATTCTCTACAGTGGAAACCAACCAATGTTTAGG 2899
750 rSerAlaGlyThrLeuAlaValSerHisAspProVal----- 762
2900 AGACATTATAAGAGTGAAGTAAAGCCAGAGTCCAACTTGTTCATCTGCACCTGTGGAAAG 2959
762 ----- 762
2960 ACTGCCGGGGAACACAGGCACGGAAGGCTTGCCCATGCTGGCCCTTTACACTAGAAGA 3019
763 -----LysProValThrThrSerProLeuProAlaGluThrAs 775
3020 TGATGTTTTTGTGGACAGCGAGGAGCGCTACTCTACTGTGTGCTGCTCTTAATAGTG 3079
775 pIleCysLeuPheSerGlySer-----IleThrGlnValGlySerProValG1 791
3080 TTTGCTGTGTGATTCAGATTCAGAGTTATGTGTAGTGTTCGCAATGTGTCTG 3139
791 ySerAlaMetAsnLeuIleProGluAsp-----GlyLeuProPheIleLeuIleSerTh 809
3140 CAACAACATTAGTCTTCTACAGACAGTACGATTTGTTGCTCTCCCTCTTCTGGCCACAT 3199
809 rGlyValLysGlyAspTyr---AlaValGluGluLysProSerGlnIleSerValMetGl 828
3200 TGGAGAAGATGGATCAGTGCCTGTTGTGCAAGAACGTCCTCTCATGAGCAGATCATTTT 3259
828 nGlnLeuGluAspGlyGlyProAspProLeuValPheValLeuAsnAlaAsnLeuLeuSe 848
3260 ACTTCTTGAAGGGGAGGATTTCTCTCTGCTACATTTGTTCTTAATGCAAACTCTTGT 3319
848 rMetValLysIleValAsnTyrValAsnArgLysCysTrpCysPheThrThrLysGlyWe 868
3320 GAATGTCAAGTTGGTGTCTTATTCCTCAGAGAAGTATTGGTACTTCTCAACCAATGGATT 3379
868 thisAlaValGlnSerGluIleValIleLeuLeuGlnCysLeuProAspGluLysCy 888
3380 GCAATGCTTGGACAGGAGGAAATATTGTTCTGTTGCAATGTTTCCAAATAGAGATAC 3439
888 sLeuProLysAspIlePheAsnHisPheValGlnLeuTyrArgAspAlaLeuAlaGlyAs 908
3440 GGTTCCTCAGGACATCTTCAGACTGTTCAATCACCATAATATAAGGATGCTTTAAAGGAAA 3499
908 nValValSerAsnLeuGlyHisSerPhePheSerGlnSerPheLeuGlySerLysGluHi 928
3500 ATACATAGAAATTTGGACAATCTGACCTTACTGAGAGTTTCTCAACAGCAGCAAGATCA 3559
928 sGlyGlyPheLeuTyrValThrSerThrThrGlnSerLeuGlnAspLeuValLeuProTh 948
3560 TGGAGGGTTCTGTTTATTACCTTACCTTCCAGAACCTTGTATGTTCTTCCCGTCCCAAG 3619


```

QY 948 rProProTyrLeuPheGlyLeuLeuIleGlnLysTrpGluThrProTrpAlaLysValPh 968
Db 3620 GAGTCCTTTCTCTCGGCATCTTATCCAGAGCTGGAGATTCCCTGGGCAAGGTTT 3679
QY 968 eProfileArgLeuMetLeuArgLeuGlyAlaGluTyrArgLeuTyrProCysProLeuPh 988
Db 3680 TCCTATCGCGTAAATGTTGCATTTGGGAGCAGATACAAACATACCTGCTCTCTAAC 3739
QY 988 eSerValArgPheArgLysProLeuPheGlyGluThrGlyHisThrIleMetAsnLeuLe 1008
Db 3740 AAGTGTACAGCAGCAGAACTCTTTTGGAGAAATAGGACATATATTATGAATCTACT 3799
QY 1008 uAlaAspPheArgLeuTyrGlnTyrThrLeuProValValGlnGlyLeuValValAspMe 1028
Db 3800 TGTGATCTCGAAATACCAAGTATACATTCGATACATTCGATCAGCTATTGATTCATAT 3859
QY 1028 tGluValArgLysThrSerIleLysIleProSerAsnArgTyrAsnGluMetLysAl 1048
Db 3860 GGAATGGGAAGAAGCTGTATAAAATACCTCGAAAAGTACAGTGTATGATGAAGGT 3919
QY 1048 aMetAsnLysSerAsnGluHisValLeuAlaGlyGlyAlaCysPheAsnGluLysAlaAs 1068
Db 3920 AATACATTCCTTAACGAGCAGCTCATCAGTATTGGAGCTAGTTTATTAGTACAGAAGCAGA 3979
QY 1068 pSerHisLeuValCysValGlnAsnAspAspGlyAsnTyrClnThrGlnAlaLysSerIl 1088
Db 3980 TTTCATCTGCTGCTGTACAGAGT---GATGGAGTTTATCAGACTCAGGCCAACAGTGC 4036
QY 1088 eHisAsnGlnProArgLysValThrGlyAlaSerPhePheValPheSerGlyAlaLeuLy 1108
Db 4037 CACTGGTTCAGCAAGCAAGTGCAGAGTGGCAAGTTTGTGTGTTTAAATGGTGTCTGAA 4096
QY 1108 sSerSerSerGlyTyrLeuAlaLysSerSerIleValGluAspGlyValMetValGlnIl 1128
Db 4097 AACATCTTCAGGATTTCTTGCATATCCAGCATAGTGGTAAATGGTAAATGGTAAAT 4156
QY 1128 eThrAlaGluAsnMetAspSerLeuArgGlnAlaLeuArgGluMetLysAspPheThrIl 1148
Db 4157 AACTCCAGAGCACCATGAGGCTTGGCGCTGCTTTACGAGAGCAAGACATTTAGAA 4216
QY 1148 eThrCysGlyLysAlaAspAlaGluGluProGlnGluHisIleHisIleGlnTrpValAs 1168
Db 4217 CCAGTGTGGAAGTGTATGCGGTAGACCTCAGAGAGTATGTGGACATCTGTTGGGTGA 4276
QY 1168 pAspAspLysAsnValSerLysGlyValValSerProLysAspGlyLysSerMetGluTh 1188
Db 4277 TTCAGAGAGGAGCAAGCAAGTGTATCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 4336
QY 1188 rIleThrAsnValLysIlePheHisGlySerGluTyrLysAlaAsnGlyLysValIleAr 1208
Db 4337 ATTTCCAAAGTGAATAAATAAACTGGAACAGATTTTGAACCTCAGAGAGACTGTGA 4396
QY 1208 gTrpThrGluValPhePheLeuGluAsnAspAspGlnHisAsnCysLeuSerAspProAl 1228
Db 4397 ATGCACCTGAGGTGTCTACTTCTTAAAGGACCAAGATATATCTATTATCA----- 4448
QY 1228 aAspHisSerArgLeuThrGluHisValAlaLysAlaPheCysLeuAlaLeuCysThrGl 1248
Db 4449 -AGTAGTACCAAGTTCCTCAAGAGGAGATGTGTGGCTGTGTAGTCTGCCCTGTGCCCTCA 4507
QY 1248 nLeuLysLeuLeuLysGlyAspGlyMetThrLysLeuGlyLeuArgValThrLeuAspse 1268
Db 4508 CCTGAGAACTCTAAAGAGTAACAGATGAACAAAATTTGGAAGTGTCCATTCAGAC 4567
QY 1268 rAspGluValGlyTyrGlnAlaGlySerAsnGlyGlnHisLeuProSerClnTyrMetAs 1288
Db 4568 TGATATGTTGAGTTCAGGAGGCTGTGAAGGCCAGCTTCTGCCCTCAGCATTTACCTGAA 4627
QY 1288 nAspPheAspSerAspLeuValLysMetIleHisGlyGlyAlaCysGlnLeuSerGluGl 1308
Db 4628 TGATCTGATAGTCCCTGATCTCCCTGATCTCCAGCGGAGCTCCCAAC---TCTAGTCT 4684
QY 1308 yProValValMetGluLeuIlePheTyrIleLeuGluAsnIle 1322

```

```

Db 4685 GCCACTAGAAATAGAGTTAGCATTTTTCATCTTAGAAAACCTT 4727

```

RESULT 14

```

AKI22241
LOCUS Mus musculus mRNA for mKIAA0305 protein. linear ROD 15-MAR-2003
DEFINITION AKI22241.1 GI:28972138
ACCESSION AKI22241
VERSION 1
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Okazaki,N., Kikuno,R., Ohara,T., Inamoto,S., Aizawa,H., Yuasa,S.,
Kakajima,D., Nagase,T., Ohara,O. and Koga,H.
TITLE Prediction of the coding sequences of mouse homologues of KIAA
gene: II. The complete nucleotide sequences of 400 mouse
KIAA-homologous cDNAs identified by screening of terminal sequences
of cDNA clones randomly sampled from size-fractionated libraries
JOURNAL DNA Res. 10, 35-48 (2003)
AUTHORS Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
TITLE Direct submission
JOURNAL Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research
Institute, Laboratory for Genome Informatics, 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT The CREATE program supported by Japan science and technology
corporation; cDNA full insert sequencing; Kazusa DNA Research
Institute; cDNA library construction, clone selection and 5'- &
3'-end one pass sequencing.
FEATURES
Location/Qualifiers
1..5438
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mbg08215"
/tissue_type="brain"
/dev_stage="adult"
/note="vector:modified pBC SK+"
gene 1..5438
/gene="mKIAA0305"
CDS 109..4719
/gene="mKIAA0305"
/note="CDS is predicted by in silico analysis. Start codon
is not identified."
/evidence=not experimental
/product="mKIAA0305 protein"
/db_xref="GI:28972139"
/translat="VLSKPTGRMDSYFKAASVGLDGLDDPEQNPKQDYLDPDAVAFN
QCVSSSEASPOLALLSKDQRCISTCASSEACEDANETFLGKTHEGLTSRPNKKN
AGLLSSVDASTDEIOPCSMRSCSKPCLISDMGLNVAHATSEEDIKOLLPPDK
SSADPLITLDSVSEALTYSVDCGSAVAREEQNNINAGINRDISIKELGVKVDMA
LFDGCLAEKAVAVANNTELVDFPMPSVLMPEQSEMSNTKDNPOYKPLFCCLLK
PLEHRTSFMKQTVKEDRNLDKNNDIYHVSDDVPPSLSCLSLSSGLCSLIHN
NEHSDILPNESEGNDVAIIBEEIQKSDVLGDETDLSKETCRSIFLPVNEKKGE
KQVEEEMVISEGLESPEASAAAGSPVLAASVPEAPGCEGTFPSSMDMGO
ELDYENIEMSGILISDAELDAFLKQCLSNSTMSAGENVDSOLQMNQITMKG
HENVAGDIYFNAEAGAGENGCGVNCETSDKENTENGLSIGEKGAITERELSCQP
DIRDLVPSTIKTQVGRKPLLSLPPGTRSRKLNKPDVDPVPESEPTANATAV
STCSADHHPDSQVFNISYDIENFEGSGSVFTANKDLSLENKRKESVLGQKQPTW
VPDEAPNMCNQVKFTFKRHRCRACQKFCVCCNRKCKQLYLEKAEKVCVYE
TINKAQAERMWSPGSGCLKSHNECATDQLOETQTSPTSPSTLPSALKQPNVE
GPCSEKQKRVFADGILPNEGAVDTAPLKSQSKCSDDFSPVLPVPMINKVDRTHS
PTVEKFNGLGDIIRSETSQSPCTHTAPVERLPGNTGEGIPMFGPTLDDVFDSE
EPSTFTVPANSGILPVASISYRLICSAKVCVNNISLLPDDDIPLPLLATSEDS
FVVQERFSEHQIILLGEGFPFATFVNANLVNVLVYSSEKYSKYSWYFSTNGLHG
QGAETILLQLPNEEDTVPKDIFRLFITIYKDALKGKYLENLDNLTFTEFLNSKDHG

```


Qy	566	r---	ProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAlaArgLysProPheTh	585
Db	2271	TCTGCCTGAATAAT	 ::: 	
Qy	585	rThrLeuGluGlyValAlaProValTrpValProAspSerGlnAlaProAsnCyMetLy	605	
Db	2301	GTTTTGGCCAGAAACAACCTCATCGGTTCCTGATCAGAGCTCCAACCTGTATGAA	2360	
Qy	605	sCyGluAlaAargPheThrPheThrLysAArgHisHisCyAArgAlaCySGlyLysVa	625	
Db	2361	CTGCCAAGTCAAAATTCACTTTACAAAACGGCAGCACCACTGCCGAGCATGTGGGAAGT	2420	
Qy	625	lPheCysAlaSerCysCySerLeuLysCySylLeuLeuTyMetAspArgLysGluAl	645	
Db	2421	ATTTTGTTGGTCTGTTTGTACAGGAATGTAGCTTCAGTAICTAGAG--AAGGAAGC	2477	
Qy	645	aArgValCyVallleCysHisSerValLeuMetAsnAlaGlnAlaTrpGluAsnMetMe	665	
Db	2478	AAGAGTATGTGTAATCTGCTATGAGACTATAAAATAAAGCTCAGGCATTTCAAAGGATGAT	2537	
Qy	665	tSerAlaSerSerClnSerProAsnProAsnAsnProAlaGluTyCySerThrIlePr	685	
Db	2538	GAGTCCAGGTGGTCTTGTCTTAAGTCTAATCTTAATGAA---TGTGCCACTGATCA	2594	
Qy	685	oProLeuGlnGlnAlaGlnAlaSerGlyAlaLeuSerSerProProThrValMetVa	705	
Db	2595	GCCACTTTCAGGAGCCCCAACATCC-----AGTACACCTTCACCTACAACCTTG--	2643	
Qy	705	lProValGlyValleuLysHisProGlyAlaGluValAlaGlnProArgGluGlnArgAr	725	
Db	2644	-CCCATCTCAGCACTGAAACACCACCAATGTTGAAGGACCATGTTCCAAAAGAACAGAGAG	2702	
Qy	725	sValTrpPheAlaAspGlyIleLeuProAsnGlyGluValAlaAspAlaAlaLysLeu--	744	
Db	2703	AGTATGTTTTCAGATGGCATCTGCGCAATGGTGAAGTTGCAGATACTACGAAATTATC	2762	
Qy	745	----	----ThrMe	746
Db	2763	ATCTGGAAGTAAAGATGTTCTGTGATCTTTAGTCTGTCTTACTGTGATGTGCCACAAT	2822	
Qy	746	t-----	-----AsnGlyThrSe	750
Db	2823	GATAAATAAGTGGATCGTACCATTCTCTACAGTGGAAAAACCAACAATGGTTTAGG	2882	
Qy	750	rSerAlaGlyThrLeuAlaValSerHisAspProVal-----	762	
Db	2883	AGACATTATAAGAAGTGAGATAGACCAAGAGTCCCACTGTGCATCTGCACCTGTGGAAG	2942	
Qy	762	-----	-----	762
Db	2943	ACTCCGGGNACACAGCCAGGAGCTTGCCCATCGCTTGCCCTTTCACACTAGAGA	3002	
Qy	763	-----LysProValThrThrSerProLeuProAlaGluThrAs	775	
Db	3003	TGATGTTTTTGTGACACGAGGAGCGGTCTACTCTCTACTGTTGTTCCTCTCTAATATG	3062	
Qy	775	pIeCysLeuPheSerGlySer-----IleThrGlnValGlySerProValGl	791	
Db	3063	TTTGCTGTTGCTAGCATTTTCAGACTACAGGTATTGTGTAGTGTGCCAAATGTGCTG	3122	
Qy	791	ySerAlaMetAsnLeuLeuProGluAsp-----GlyLeuProProThrLeuLeuSerTh	809	
Db	3123	CAACAACATTAGTCTTCTACGACGATGACATTGGTTTGCCTCCCCTTCTGCGCCATC	3182	
Qy	809	rGlyValLysGlyAspTyr--AlaValGluGluLysProSerClnIleSerValMetGl	828	
Db	3183	TGGAGAAGATGATCAGTGCTGTTGTGTCAGAAACGTCCTCTCATGAGCAGATCATTT	3242	
Qy	828	nGlnLeuGluAspGlyGlyProAppProLeuValPheValLeuAsnAlaAsnLeuLeuSe	848	
Db	3243	ACTTCTTGAAGGGGAAGGATTTCTCTCTGCTACATTTGTTCTAAATGCAAACTCTTGT	3302	

Qy	848	rMetVallyIleValAsnTyrValAsnArgLysCysTyrCysPheThrThrLysGlyMe	868
Db	3303	GAATGTCAAGTCTGGTGTCTTTATTCCTCAGAGAAGTATTGGTACTCTCAACCAATGGATT	3362
Qy	868	tHisAlaValGlnSerGluIleValIleLeuLeuGlnCysLeuProAspGluLysCy	888
Db	3363	GCATGGCTTGGGACAGCAGCAAAATTATGTCCTTGGCAATGTTTCCCAATGAGGATAC	3422
Qy	888	sLeuProLysAspIlePheAsnHisPheValGlnLeuTyrArgAspAlaLeuAlaGlyAs	908
Db	3423	GGTTCCTTAAGACATCTTTCAGATGTGTTCATPCACCATATATAAGGATGCTTTAAAGGAAA	3482
Qy	908	nValValSerAsnLeuGlyHisSerPhePheSerGlnSerPheLeuGlySerLysGluHi	928
Db	3483	ATACATAGAAAAATTGGACAATCTGACCTTTACTGAGAGTTTTCTCAACGCAAGATCA	3542
Qy	928	sGlyGlyPheLeuTyrValThrSerThrTyrGlnSerLeuGlnAspLeuValLeuProTh	948
Db	3543	TGGAGGGTCTCTGTATTATACACTCTCCAGAACCTTGATGGTCTCCCCGTGCCAAG	3602
Qy	948	rProProTyrLeuPheGlyIleLeuIleGlnLysTyrPGLuThrProTyrAlaLysValPh	968
Db	3603	GAGTCCCTTTTCTCTGGGCAATCTTATCCAGAAGCTGGAGATTCCTCTGGCAAGGTTTT	3662
Qy	968	eProIleArgLeuMetLeuArgLeuGlyAlaGluTyrArgLeuTyrProCysProLeuPh	988
Db	3663	TCCTATCGGCTAATGTTGGATGGGAGCAGAAATACAAAGCATACCTCTCTCTAAC	3722
Qy	988	eSerValArgPheArgLysProLeuPheGlyGluThrGlyHisThrIleMetAsnLeuLe	1008
Db	3723	AAAGTGTCCAGAGGCGAGAAACCTCTTTTGAGAAATAGGACATACATTATATGAACATTACT	3782
Qy	1008	uAlaAspPheArgAsnTyrGlnTyrThrLeuProValValGlnGlyLeuValValAspMe	1028
Db	3783	TGTTGATCTTCGAAATTACCAGTATACATTGCATATACATCCGATCAGCTATTGATTCATAT	3842
Qy	1028	tGluValArgLysThrSerIleLysIleProSerAsnArgTyrAsnGluMetMetLysAl	1048
Db	3843	GGAAATCGGAAAAAGCTGTATAAAATACCTCGAAAAAGTACAGTGATGTATGAAGGT	3902
Qy	1048	aMetAsnLysSerAsnGluHisValLeuAlaGlyGlyAlaCysPheAsnGluLysAlaAs	1068
Db	3903	AATACATCTCTTAACGAGCAGCTATCATGATTGGAGCTAGTTTGTAGTACAGAAGCAGA	3962
Qy	1068	pSerHisLeuValCysValGlnAsnAspAspGlyAsnTyrGlnThrGlnAlaIleSerIl	1088
Db	3963	TTCTCATCTGGTCTGTGTACAGAGT--GATGAGTTTATCAGACTCAGGCCAACAGTGC	4019
Qy	1088	eHisAsnGluProArgLysValThrGlyAlaSerPhePheValPheSerGlyAlaLeuLy	1108
Db	4020	CACTGGTGCACCCAAAGAAAGTACAGGTGCAAGTTTGTGGTGTATTATGTTGGTCTCTGAA	4079
Qy	1108	sSerSerSerGlyTyrLeuAlaLysSerSerIleValGluAspGlyValMetValGlnIl	1128
Db	4080	AACATCTTCAGGATTCTTCCCAATCCAGCATAGTCGAAGATGGCTTAATGTGCAAAAT	4139
Qy	1128	eThrAlaGluAsnMetAspSerLeuArgGlnAlaLeuArgGluMetLysAspPheThrIl	1148
Db	4140	AACTCCAGACACCATGGAGGGCTTGGCGTGGCTTTACGAGACAAAGACATTTAGAAT	4199
Qy	1148	eThrCysGlyLysAlaAspAlaGluGluProGlnGluHisIleHisIleGlnTyrValAs	1168
Db	4200	CCAGTGTGAAAAGTTGATCCGTTAGACCTGAGAGAGTATGTGGACATCTGTGTGGTGGGA	4259
Qy	1168	pAspAspLysAsnValSerLysGlyValValSerProIleAspGlyLysSerMetGluTh	1188
Db	4260	TTCAGAAAGAGAGGAACAAAGAGTGTTATCAGTTCAGTAGATGGATGTTCAGTGGAAAGG	4319
Qy	1188	rIleThrAsnValLysIlePheHisGlySerGluTyrLysAlaAsnGlyLysValIleAr	1208
Db	4320	ATTTCCAAAGTGAAAAATAAACTGGAAACAGATTTTGAAACTGAGGAGAGACTGTGAA	4379
Qy	1208	gTyrThrGluValPhePheLeuGluAsnAspAspGlnHisAsnCysLeuSerAspProAl	1228

QY 141 ---ProSerGln-----LeuIysAspAspGlySerIleGlyArgAspProSer 155
Db 922 CACAAAGCCCGCATGGGATTACTAAAGATCTTGCTTAGTAAAGAGGAGTAGAT 981
QY 156 MetSerAlaIleThrSerLeuThrValAspSerValIleSerSerGlnGlyThrAspGly 175
Db 982 GTGGCAGTCATACTGCGCA-----GAATGTTTAAAGAGAGGCAAGACAAGT 1032
QY 176 CysProAlaValLysLysGlnGluAsnTyrIleProAspGluAspLeuThrGlyLysIle 195
Db 1033 GCTTTGACCTGCAGCCCTCCGAAAAT-----GAAGATTTATGCTTAAATGAT 1080
QY 196 SerSerProArgThrAspLeuGlySerProAsnSerPheSerHisMetSerGluGlyIle 215
Db 1081 TCAATTCAGAGATGAATTAATTCAAATTTACCTGACTTTTCCCTTCAGGAAGATPAAGACT 1140
QY 216 LeuMetLysLysGluProAlaGluGluSerThr----- 227
Db 1141 GTTATAAAACAATCTGCACAAAGAGACTCAAAAAGTTTAGACCTTAAAGGATAATGATGA 1200
QY 228 GluGluSerLeuArgSerGlyLeuProLeuLeuLeuLysProAspMetProAsnGlySer 247
Db 1201 ATCCAGATTCCTCTCAGCTTTACATGTTTCCAGTAAA---GATGCGCGTCTCTCATTTG 1257
QY 248 GlyArgAsnAsnAspCysGluArgCysSerAspCysLeuValProAsnGluValArgAla 267
Db 1258 TCCGTGCTCTCGGTCTGGGTCTATGAGTGGATCATTAATGAAAGTAAGACCGGGT 1317
QY 268 AspGluAsnGluGlyTyrGluHisGluGluThrLeuGlyThr----- 282
Db 1318 GATTTTTCATCAGATGAACATAAAGATAATACAAAGATGACAGTGCAGTACATACATGAA 1377
QY 283 GluPheLeuAsnMetThr-----GluHisPheSerGluSerGlnAspMetThr 298
Db 1378 GAAATACAGACAGCTGTGTTCTAGGTGGGAACCATTCAAAGAGAATGATCTTTTGAAA 1437
QY 299 AsnTrpLys-----LeuThrLysLeuAsnGlu---MetAsnAspSerGln 312
Db 1438 CAGGAAAAATGTAAAGACATCTCTTCAGTCATTAATTTGAAGGGATGGAAGACAGAAAG 1497
QY 313 ValAsnGluGluLysGluLysPheLeuGlnIleSerGlnProGluAspThrAsnGlyAsp 332
Db 1498 ATAGATCCTGACCCAG-----ACAGTAAATCAGAGCTGAGTCTTTGGATGGTGGT 1545
QY 333 SerGlyGlyGlnCysValGlyLeuAlaAspAlaGlyLeuAspLeuLysGlyThrCysIle 352
Db 1546 GACACCAAGTTCTACAGTTGATGATCTCAAGAGGG-----CTTTCTGGCACTCATGTC 1599
QY 353 SerGluSerGluGluCysAspPheSerThrValIleAspThrProAlaAlaAsnTyrLeu 372
Db 1600 CCAGAGTCTTCTGATGTG---TGGAAGGTTTTTATTATATCTTTTCAAGCAATGATG 1656
QY 373 SerAsnGlyCysAspSerTyrGlyMetGlnAspProGlyValSerPheValProLysThr 392
Db 1657 GATGGGCAAGCTTAGATTACTTTAATTTGATGAAGCGCAAAAGTGCCCACTAATT 1716
QY 393 LeuProSerLysGluAspSer---ValThrGlu----- 402
Db 1717 AGTGATGCTGAACCTGATGCTTCTTCAGCAGAACGATCTCTCAGACCACTAATAAAG 1776
QY 403 -----GluLysGluIleGluGluSerLysSerGluCys----- 413
Db 1777 TCTTTGAAGAAATGAATGACTCTAATTCGGAATGAATGAATCAGATAGATATGAAGGC 1836
QY 414 -----TyrSerAsnIleTyr---GluGlnArgGlyAsnGluAlaThrGlu 427
Db 1837 TTAGATGATGGAACATCAATAATATATTTTCAATCAGAACGAGGAGCTATTTGGGAA 1896
QY 428 GlySerGlyLeu---LeuLeuAsnSerThrGlyAspLeuMetLysLysAsnTyrLeuHis 446
Db 1897 AGTCATGCTATTAATAATTTGTGAACACAGTTGAT-----AAACAAAATACATAGAA 1950

QY 447 AsnPheCys-----SerGlnValProSerValLeuGlyGlnSerSerPro 461
Db 1951 AATGGCCCTTCTTTAGGAGAAAAACACTATTCCAGTTTCAAAAGGGTTACTTACCAGT 2010
QY 462 Lys-----ValValAlaSerLeuProSerIleSerValProPhe 474
Db 2011 AAGTCTGAGATTACAATCAATATCATCTCTCTGATATTAAACAGTCAATCTGTT----- 2064
QY 475 GlyGlyAlaArgProLysGlnProSerAsnLeuLysLeuGlnIleProLysProLeuSer 494
Db 2065 GGAGGGGCCAGACCTTAAGCAATTTGTTAGCCTT-----CCATCAAGA 2106
QY 495 AspHisLeuGlnAsnAspPheProAlaAsnSerGlyAsnAsnThrLysAsnLysAsnAsp 514
Db 2107 ACAAGGAGTTCAAGACCTG-----AATAAGCCAGAT 2139
QY 515 IleLeuGlyLysAlaLysLeuGlyGluAsnSerAlaThrAsnVal----- 529
Db 2140 GTTCCAGATACAAATAGAAAGTGAACCCAGCAGCAGATACCGTTGTTTCCAATCACTTGT 2199
QY 530 -----CysSerProSerLeu-----GlyAsnIleSerAsnValAsp 541
Db 2200 GCTATAGATCTACAGCTGATCCAGGTTAGCTTCAACTCTAATTAATGATATAGAA 2259
QY 542 ThrAsnGlyGluHisLeuGluSerTyrGluAlaGluIleSerThrArgProCysLeuAla 561
Db 2260 AGTAATTTCTGAAGTGGATCTAGTTTC-----GTAACT 2292
QY 562 LeuAlaProAspSer---ProAspAsnAspLeuArgAlaGlyGlnPheGlyIleSerAla 580
Db 2293 GCAATAGAGATTTCTGTACCTGAAACACTTGCAAAGAGAGGC----- 2334
QY 581 ArgLysProPheThrThrLeuGlyGluValAlaProValTrrValProAspSerGlnAla 600
Db 2335 -----TTGGTTTGGCCAGAAACAGCCCTACTTGGGTTCTTGATTCAGAACT 2382
QY 601 ProAsnCysMetLysCysGluAlaArgPheThrPheThrLysArgArgHisCysArg 620
Db 2383 CCAAACTGTATGAATGCGCAAGTCAATTTACTTTTACAAACGCGCAGACCATTCGCGA 2442
QY 621 AlaCysGlyLysValPheCysAlaSerCysCysSerLeuLysCysLysLeuTyrMet 640
Db 2443 GCATGTGGGAAAGTATTTTGTGCTGTCTGTATAGGAAGTGTAAACTGCAATATCTTA 2502
QY 641 AspArgLysGluAlaArgValCysValIleCysHisSerValLeuMetAsnAlaGlnAla 660
Db 2503 GAA---AAGGAAGCAAGAGTATGTCTAGTCTCTATGAAACTATTAGTAAAGCTCAGGCA 2559
QY 661 TrpGluAsnMetMetSerAlaSerSerGlnSerProAsnProAsnAsnProAlaGluTyr 680
Db 2560 TTTGAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2616
QY 681 CysSerThrIleProLeuGlnGlnAlaGlnAlaSerGlyAlaLeuSerSerProPro 700
Db 2617 TGTACTACTGTCCAGCTCCTCAGGAGAACCAACATCC-----AGTATACCTTCA 2667
QY 701 ProThrValMetValProValGlyValLeuLysHisProGlyAlaGluValAlaGlnPro 720
Db 2668 CCAGCAACTTTG---CCAGTCTCAGCACTTAAACCAACCCAGGTTTGAAGGAGCTATGTTCC 2724
QY 721 AsgGluGlnArgValTrrPheAlaAspGlyIleLeuProAsnGlyGluValAlaAsp 740
Db 2725 AAAGAACAGAGAGATGTTGTTGCGATGATGATGATGATGATGATGATGATGATGATGAT 2784
QY 741 AlaAlaLysLeuThrMetAsnGlyThrSerSerAlaGlyThrLeuAlaValSerHisAsp 760
Db 2785 ACAACAAAATTA-----TCATCTGGAAGTAAAGAGATGTTCTGAAGAC 2826
QY 761 -----ProValLysProValThrThr----- 767
Db 2827 TTTAGTCTCTCTCCTGATGCTGCTATGACAGTAAACACAGTGGATCATTTCCCATCTCT 2886
QY 767 ----- 767

Db	2987	ACTACAGTGGAAAGCCAAACAATGACAGGAGATATTACAGAAATGAGATAATTACG	2946
Qy	768	-----SerProLeuProAlaGluThrAspIleCysLeu-----	778
Db	2947	AGTCCTATTCTCAGGTTCTCATCGAGGAAAAATTGCTATGATGAACACAGAAATGAGGG	3006
Qy	779	-----PheSerGlySerIleThr-----GlnVal	786
Db	3007	TTACCTACTTCTGGTTCATTTACACATAGATGATGTTTTCGAGAAACTGAAGAACCA	3066
Qy	787	GlySerProValGlySerAlaMetAsn-----	795
Db	3067	TCTAGTCTCTAGTGTCTTAGTTAAACAGCAATTTACCTATTGCTAGTATTTCAGATTAT	3126
Qy	796	-----LeuIlePro-----	798
Db	3127	AGGTTACTGTGTGATATTAAACAAGTATGTCGCAATAAGATTAGTCTCTACCTAATGAT	3186
Qy	799	GluAspGlyLeuProProlIleLeuIleSerThrGlyValIleGlyAspTyr---AlaVal	817
Db	3187	GAGACAGATTGCCCCCACTCTGGTTGCTATCGAGAAAAAGGATCAGTGCCTGTAGTA	3246
Qy	818	GluGluIlyProSerGlnIleSerValMetGlnGlnLeuGluAspGlyGlyProAspPro	837
Db	3247	GAGAACATCCATCTCATGACGACATCATTTTGCCTTCTTGAAGGTGAAGCTTTCATCCT	3306
Qy	838	LeuValPheValLeuAsnAlaAsnLeuLeuSerMetValIlySileValAsnTyrValAsn	857
Db	3307	GTTCATATTGTCCTAAATGCTAATCTACTCGTGAATGCTCAAAATCATATTTTATTCTCTCA	3366
Qy	858	ArgIlyCysTrpCysPheThrLysGlyMetHisAlaValGlnGlnSerGluIleVal	877
Db	3367	GACAAATATTGGTACTTTTCAACAATGGAATGCTATGCTTGGCAGCAGGCAAAATATT	3426
Qy	878	IleLeuLeuGlnCysLeuProAspGluIlyCysLeuProIlyAspIlePheAsnHisPhe	897
Db	3427	ATTCTATTGTTATGTTTGCCAAATGAAGATACTATTCTTAAGACATCTTCACATATT	3486
Qy	898	ValGlnLeuTyrArgAspAlaLeuAlaGlyAsnValValSerAsnLeuGlyHisSerPhe	917
Db	3487	ATCACCATATATAGGATGCTCTAAAGGAAAAATACATAGAAACTTGGACAATATTACC	3546
Qy	918	PheSerGlnSerPheLeuGlySerLysGluHisGlyGlyPheLeuTyrValThrSerThr	937
Db	3547	TTACTGAGAGTTTCTCAGTAGCAAGATCATCGGAGGATTCCTGTTTATTACACCTACT	3606
Qy	938	TyrGlnSerLeuGlnAspLeuValLeuProThrProProTyrLeuPheGlyIleLeuIle	957
Db	3607	TTTCAGAAACTTGTATGATCTCTCATTAACCAAGTAATCTTTCTTTGTGGAATTCATC	3666
Qy	958	GlnLysTrpGluThrProTlpAlaLysValPheProIleArgLeuMetLeuAsgLeuGly	977
Db	3667	CAGAACTTGAGATTCCTGGCCAAAGTTTTTCTATGCTTTTAAATGTTGAGATTGGGT	3726
Qy	978	AlaGluTyrArgLeuTyrProCysProLeuPheSerValArgPheArgLysProLeuPhe	997
Db	3727	GCAGATATAAAGCATATCTGCTCTCTCTAACAGCATCAGAGCCGCAAAACCTCTTTT	3786
Qy	998	GlyGluThrGlyHisThrIleMetAsnLeuLeuAlaAspPheArgAsnTyrGlnTyrThr	1017
Db	3787	GGAGAAATAGACACATATTATGAACTTTACTTGTTCACCTTGAAATTCACAGTATACC	3846
Qy	1018	LeuProValValGlnGlyLeuValValAspVetGluValArgIlyThrSerIleLysIle	1037
Db	3847	TTGCATAATATAGATCAACTGTTCATTCATATGAAATGCAAAAGGCTGCATAAAATA	3906
Qy	1038	ProSerAsnArgTyrAsnGluMetMetLysAlaMetAsnLysSerAsnGluHisValLeu	1057
Db	3907	CCACGAAAAAGTACAGTGATGTAATGAAGTACTAAATTCCTCCAAATGACATGTCAIT	3966
Qy	1058	AlaGlyGlyAlaCysPheAsnGluAlaAspSerHisLeuValCysValGlnAsnAsp	1077

3967	AGCA	NTGGAGCAAGTTTCAGTACAGAAAGCAGATTCTCATCTAGTCTGTATACAGAAT---	4023
1078	AspGly	AsnTyrGlnThrGlnAlaIleSerIleHisAsnGlnProArgLysValThrGly	1097
4024	GATGGA	ATTATGAACACACAGGCCAACAGTACCACCTGGCCATCTCTAGAAAAGTGACAGGT	4083
1098	AlaSer	PhePheValPheSerGlyAlaLeuLysSerSerSerClyTyrLeuAlaLysSer	1117
4084	GCAAG	TTTTGTGTATTCAATGGAGCTCTAAAACAATCTTCAGGATTTCTTGTAAAGTCC	4143
1118	SerIle	ValGluAspGlyValMetValGlnIleThrAlaGluAsnMetAspSerLeuArg	1137
4144	AGCAT	AGTTTGAAGATGGCTTATATGGTACAAATAACTCCAGAGACCATGAATGGCTTGGCG	4203
1138	GlnAla	LeuArgGluMetLysAspPheThrIleThrCysGlyLysAlaAspAlaGluGlu	1157
4204	CTAGC	TTTACGGAACACAGAAAGACTTTAAATTTACATGTGGGAAAGTTGATGCAGTAGAC	4263
1158	ProGln	GluHisIleHisIleGlnThrValAspAspLysAsnValSerLysGlyVal	1177
4264	CTGAGA	GAATACATCGTATATCTCTCGGTAGATGCTTGAAAGAAAAGCAACAAGAGATT	4323
1178	ValSer	ProIleAspGlyLysSerMetGluThrIleThrAsnValLysIlePheHisGly	1197
4324	ATCAG	TTTCAGTGGATGGAATATCATTTACAGGATTTCCAAGTGAAGAAAATAAAAC	4383
1198	SerGlu	TyrLysAlaAsnGlyLysValIleArgTyrThrGluValPhePheLeuGluAsn	1217
4384	GCAGA	TTTTGAACCGATGAGAAGATTGTAATAATGTACCGAGGTGTCTACTTTCTAAAG	4443
1218	AspAsp	GlnHisAsnCysLysSerAspProAlaAspHisSerArgLeuThrGluHisVal	1237
4444	GACC	CAGGATTTATCTATTATTATCA-----ACTCTTATCAGTTGTCAAAGAAATA	4494
1238	AlaLys	AlaPheCysLeuAlaLeuCysThrGlnLeuLysLeuLeuLysGlyAspGlyMet	1257
4495	GCCAT	GGCTGTGTAGTCTGGCTGTGCCCTCACCTGAAACCTCTAAAGATAATGGCATG	4554
1258	ThrLys	LeuGlyLeuArgValThrLeuAspSerAspGlnValGlyTyrGlnAlaGlySer	1277
4555	AATAAA	TTTGGACTCAGAGTTTCCATGTGACACTGATGTGTGAATTTCTCAGGCAGGATCT	4614
1278	AsnGly	GlnHisLeuProSerGlnTyrMetAsnAspPheAspSerAspLeuValLysMet	1297
4615	GAAGG	CCAACTCTGCCCTCAGCATATCTAAATGATCTTGATAGTCTCTGATACCTGTG	4674
1298	IleHis	GlyGlyAlaCysGlnLeuSerGluGlyProValValMetGluLeuIlePheTyr	1317
4675	ATCCT	AGTTGGGACCTTCCAAAC---TCTAGTTTACCATTAGAAATAGAAATAGTGTTC	4731
1318	IleLeu	GluAsnIle 1322	
4732	ATTAT	AGACATCTT 4746	

Search completed: May 1, 2004, 01:51:55
Job time : 11601.7 secs